(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

CORRECTED VERSION

10/009384

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
9 November 2000 (09.11.2000)

PCT

(10) International Publication Number
WO 00/66143 A1

(51) International Patent Classification[7]: A61K 38/00, 39/00, 39/02, 39/04, C07H 21/02, 21/04, C07K 1/00, 14/00, C12N 1/12, 1/20, C12P 19/34, 21/04, 21/06

(21) International Application Number: PCT/US00/12197

(22) International Filing Date: 4 May 2000 (04.05.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

| | | |
|---|---|---|
| 60/132,479 | 4 May 1999 (04.05.1999) | US |
| 60/132,503 | 4 May 1999 (04.05.1999) | US |

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

| | |
|---|---|
| US | 60/132,503 (CON) |
| Filed on | 4 May 1999 (04.05.1999) |
| US | 60/132,479 (CON) |
| Filed on | 4 May 1999 (04.05.1999) |

(71) Applicant *(for all designated States except US)*: **THE PUBLIC HEALTH RESEARCH INSTITUTE OF THE CITY OF NEW YORK, INC.** [US/US]; 455 First Avenue, New York, NY 10016 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **GENNARO, Maria, L.** [US/US]; 25 Central Park West, Apt. 11-T, New York, NY 10023 (US). **GOMEZ, Manuel, J.** [US/US]; 55 River Drive South, Apt. 204, Jersey City, NJ 07310 (US).

(74) Agent: **HONE, William, J.**; Fish & Richardson P.C., Suite 2800, 45 Rockefeller Plaza, New York, NY 10111 (US).

(81) Designated States *(national)*: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published:**
*with international search report*

(48) Date of publication of this corrected version:
14 February 2002

(15) Information about Corrections:
see PCT Gazette No. 07/2002 of 14 February 2002, Section II
**Previous Correction:**
see PCT Gazette No. 14/2001 of 5 April 2001, Section II

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

WO 00/66143 A1

(54) Title: SECRETED PROTEINS OF *MYCOBACTERIUM TUBERCULOSIS* AND THEIR USE AS VACCINES AND DIAGNOSTIC REAGENTS

(57) Abstract: The invention provides *mycobacterium tuberculosis* polypeptides and genes encoding them for use in diagnostic and prophylactic methodologies.

# SECRETED PROTEINS OF *MYCOBACTERIUM TUBERCULOSIS* AND THEIR USE AS VACCINES AND DIAGNOSTIC REAGENTS

## Background of the Invention

The invention is in the field of tuberculosis and, specifically, reagents useful for generating immune responses to *Mycobacterium tuberculosis* and for diagnosing infection and disease in a subject that has been exposed to *M. tuberculosis*.

Tuberculosis infection continues to be a world-wide health problem. This situation has recently been greatly exacerbated by the emergence of multi-drug resistant strains of *M. tuberculosis* and the international AIDS epidemic. It has thus become increasingly important that effective vaccines against and reliable diagnostic reagents for *M. tuberculosis* be produced.

U.S. application no. 08/796,792 is incorporated herein by reference in it entirety.

## Summary of the Invention

The invention is based on the discovery of a novel group of open reading frames (ORFs) encoding polypeptides that are secreted by *M. tuberculosis*. The invention features these polypeptides, functional segments thereof, DNA molecules encoding either the polypeptides or the functional segments, vectors containing the DNA molecules, cells transformed by the vectors, compositions containing one or more of any of the above polypeptides, functional segments, or DNA molecules, and a variety of diagnostic, therapeutic, and prophylactic (vaccine) methodologies utilizing the foregoing.

Specifically, the invention features an isolated DNA molecule containing a DNA sequence encoding a polypeptide

with a first amino acid sequence that can be the amino acid sequence of the polypeptide MTSP1, MTSP2, MTSP3, MTSP4, MTSP5, MTSP6, MTSP7, MTSP8, MTSP9, MTSP10, MTSP11, MTSP12, MTSP13, MTSP14, MTSP15, MTSP16, MTSP17, MTSP18, MTSP19, MTSP20, MTSP21, MTSP22, MTSP23, MTSP24, MTSP25, MTSP26, MTSP27, MTSP28, MTSP29, MTSP30, MTSP31, MTSP32,MTSP33, MTSP34, MTSP35, MTSP36, MTSP37, MTSP38, MTSP39, MTSP40, MTSP41, MTSP42, MTSP43, MTSP44, MTSP45, MTSP46, or MTSP47, as depicted in Fig. 1, or a second amino acid sequence identical to the first amino acid sequence with conservative substitutions; the polypeptide has *Mycobacterium tuberculosis* specific antigenic and immunogenic properties. Also included in the invention is an isolated portion of the above DNA molecule. The portion of the DNA molecule encodes a segment of the polypeptide shorter than the full-length polypeptide, and the segment has *Mycobacterium tuberculosis* specific antigenic and immunogenic properties. Other embodiments of the invention are vectors containing the above DNA molecules and transcriptional and translational regulatory sequences operationally linked to the DNA sequence, the regulatory sequences allow for the expression of the polypeptide or functional segment encoded by the DNA sequence in a cell. The invention encompasses cells (e.g., eukaryotic and prokaryotic cells) transformed with the above vectors.

The invention encompasses compositions containing any of the above vectors and a pharmaceutically acceptable diluent or filler. Other compositions to be used as DNA vaccines can contain at least two (e.g., three, four, five, six, seven, eight, nine, then, twelve, fifteen or twenty) DNA sequences, each encoding a polypeptide of the *Mycobacterium tuberculosis* complex or a functional segment thereof, with the DNA sequences being operationally linked to transcriptional and

translational regulatory sequences which allow for expression of each of the polypeptides in a cell of a vertebrate. In such compositions, at least one of the DNA sequences contains the sequence of the above DNA molecules of the invention.

The invention also features an isolated polypeptide with a first amino acid sequence that can be the sequence of the polypeptide MTSP1, MTSP2, MTSP3, MTSP4, MTSP5, MTSP6, MTSP7, MTSP8, MTSP9, MTSP10, MTSP11, MTSP12, MTSP13, MTSP14, MTSP15, MTSP16, MTSP17, MTSP18, MTSP19, MTSP20, MTSP21, MTSP22, MTSP23, MTSP24, MTSP25, MTSP26, MTSP27, MTSP28, MTSP29, MTSP30, MTSP31, MTSP32, MTSP33, MTSP34, MTSP35, MTSP36, MTSP37, MTSP38, MTSP39, MTSP40, MTSP41, MTSP42, MTSP43, MTSP44, MTSP45, MTSP46, or MTSP47, as depicted in Fig. 1, or a second amino acid sequence identical to the first amino acid sequence with conservative substitutions. The polypeptide has *Mycobacterium tuberculosis* specific antigenic and immunogenic properties. Also included in the invention is an isolated segment of this polypeptide, the segment being shorter than the full-length polypeptide and having *Mycobacterium tuberculosis* specific antigenic and immunogenic properties. Other embodiments are compositions containing the polypeptide, or functional segment, and a pharmaceutically acceptable diluent or filler. Compositions of the invention can also contain at least two (e.g., three four, five, six, seven, eight, nine, ten, twelve, fifteen, or twenty) polypeptides of the *Mycobacterium tuberculosis* complex, or functional segments thereof, with at least one of the at least two polypeptides having the sequence of one of the above described polypeptides of the invention.

The invention also features methods of diagnosis. One embodiment is a method involving: (a) administration of one of the above polypeptide compositions to a subject suspected

of having or being susceptible to *Mycobacterium tuberculosis* infection; and (b) detecting an immune response in the subject to the composition, as an indication that the subject has or is susceptible to *Mycobacterium tuberculosis* infection. Another embodiment is a method that involves: (a) providing a population of cells containing CD4 T lymphocytes from a subject; (b) providing a population of cells containing antigen presenting cells (APC) expressing a major histocompatibility complex (MHC) class II molecule expressed by the subject; (c) contacting the CD4 lymphocytes of (a) with the APC of (b) in the presence of one or more of the polypeptides, functional segments, and or polypeptide compositions of the invention; and (d) determining the ability of the CD4 lymphocytes to respond to the polypeptide, as an indication that the subject has or is susceptible to *Mycobacterium tuberculosis* infection. Another diagnostic method of the invention involves: (a) contacting a polypeptide, a functional segment, or a polypeptide/functional segment composition of the invention with a bodily fluid of a subject; (b) detecting the presence of binding of antibody to the polypeptide, functional segment, or polypeptide/functional segment composition, as an indication that the subject has or is susceptible to *Mycobacterium tuberculosis* infection.

Also encompassed by the invention are methods of vaccination. These methods involve administration of any of the above polypeptides, functional segments, or DNA compositions to a subject. The compositions can be administered alone or with one or more of the other compositions.

As used herein, an "isolated DNA molecule" is a DNA which is one or both of: not immediately contiguous with one

or both of the coding sequences with which it is immediately contiguous (i.e., one at the 5' end and one at the 3' end) in the naturally-occurring genome of the organism from which the DNA is derived; or which is substantially free of DNA sequence with which it occurs in the organism from which the DNA is derived. The term includes, for example, a recombinant DNA which incorporated into a vector, e.g., into an autonomously replicating plasmid or virus, or into the genomic DNA of a prokaryote or eukaryote, or which exists as a separate molecule (e.g., a cDNA or a genomic fragment produced by PCR or restriction endonuclease treatment) independent of other DNA sequences. Isolated DNA also includes a recombinant DNA which is part of a hybrid DNA encoding additional *M. tuberculosis* polypeptide sequences.

"DNA molecules" include cDNA, genomic DNA, and synthetic (e.g., chemically synthesized) DNA. Where single-stranded, the DNA molecule may be a sense strand or an antisense strand.

An "isolated polypeptide" of the invention is a polypeptide which either has no naturally-occurring counterpart, or has been separated or purified from components which naturally accompany it, e.g., in *M. tuberculosis* bacteria. Typically, the polypeptide is considered "isolated" when it is at least 70%, by dry weight, free from the proteins and naturally-occurring organic molecules with which it is naturally associated. Preferably, a preparation of a polypeptide of the invention is at least 80%, more preferably at least 90%, and most preferably at least 99%, by dry weight, the peptide of the invention. Since a polypeptide that is chemically synthesized is, by its nature, separated from the components that naturally accompany it, the synthetic polypeptide is "isolated."

An isolated polypeptide of the invention can be obtained, for example, by extraction from a natural source (e.g., *M. tuberculosis* bacteria); by expression of a recombinant nucleic acid encoding the polypeptide; or by chemical synthesis. A polypeptide that is produced in a cellular system different from the source from which it naturally originates is "isolated," because it will be separated from components which naturally accompany it. The extent of isolation or purity can be measured by any appropriate method, e.g., column chromatography, polyacrylamide gel electrophoresis, or HPLC analysis.

The polypeptides may contain a primary amino acid sequence that has been modified from those disclosed herein. Preferably these modifications consist of conservative amino acid substitutions. Conservative substitutions typically include substitutions within the following groups: glycine and alanine; valine, isoleucine, and leucine; aspartic acid and glutamic acid; asparagine and glutamine; serine and threonine; lysine and arginine; and phenylalanine and tyrosine.

The terms "protein" and "polypeptide" are used herein to describe any chain of amino acids, regardless of length or post-translational modification (for example, glycosylation or phosphorylation). Thus, the term "*Mycobacterium tuberculosis* polypeptide" includes full-length, naturally occurring *Mycobacterium tuberculosis* protein, as well a recombinantly or synthetically produced polypeptide that corresponds to a full-length naturally occurring *Mycobacterium tuberculosis* protein or to particular domains or portions of a naturally occurring protein. The term also encompasses a mature *Mycobacterium tuberculosis* polypeptide

which has an added amino-terminal methionine (useful for expression in prokaryotic cells).

As used herein, "immunogenic" means capable of activating a primary or memory immune response. Immune responses include responses of CD4+ and CD8+ T lymphocytes and B-lymphocytes. In the case of T lymphocytes, such responses can be proliferative, and/or cytokine (e.g., interleukin(IL)-2, IL-3, IL-4, IL-5, IL-6, IL-12, IL-13, IL-15, tumor necrosis factor-α (TNF-α), or interferon-γ (IFN-γ))-producing, or they can result in generation of cytotoxic T-lymphocytes (CTL). B-lymphocyte responses can be those resulting in antibody production by the responding B lymphocytes.

As used herein, "antigenic" means capable of being recognized by either antibody molecules or antigen-specific T cell receptors (TCR) on activated effector T cells (e.g., cytokine-producing T cells or CTL).

Thus, polypeptides that have "*Mycobacterium tuberculosis* specific antigenic properties" are polypeptides that: (a) can be recognized by and bind to antibodies elicited in response to *Mycobacterium tuberculosis* organisms or wild-type *Mycobacterium tuberculosis* molecules (e.g., polypeptides); or (b) contain subsequences which, subsequent to processing of the polypeptide by appropriate antigen presenting cells (APC) and bound to appropriate major histocompatibility complex (MHC) molecules, are recognized by and bind to TCR on effector T cells elicited in response to *Mycobacterium tuberculosis* organisms or wild-type *Mycobacterium tuberculosis* molecules (e.g., polypeptides).

As used herein, polypeptides that have "*Mycobacterium tuberculosis* specific immunogenic properties" are polypeptides that: (a) can elicit the production of

antibodies that recognize and bind to *Mycobacterium tuberculosis* organisms or wild-type *Mycobacterium tuberculosis* molecules (e.g., polypeptides); or (b) contain subsequences which, subsequent to processing of the polypeptide by appropriate antigen presenting cells (APC) and bound to appropriate major histocompatibility complex (MHC) molecules on the surface of the APC, activate T cells with TCR that recognize and bind to peptide fragments derived by processing by APC of *Mycobacterium tuberculosis* organisms or wild-type *Mycobacterium tuberculosis* molecules (e.g., polypeptides) and bound to MHC molecules on the surface of the APC. The immune responses elicited in response to the immunogenic polypeptides are preferably protective. As used herein, "protective" means preventing establishment of an infection or onset of a disease or lessening the severity of a disease existing in a subject. "Preventing" can include delaying onset, as well as partiallly or completely blocking progress of the disease.

As used herein, a "functional segment of a *Mycobacterium tuberculosis* polypeptide" is a segment of the polypeptide that has *Mycobacterium tuberculosis* specific antigenic and immunogenic properties.

Where a polypeptide, functional segment of a polypeptide, or a mixture of polypeptides and/or functional segments have been administered (e.g., by intradermal injection) to a subject for the purpose of testing for a *M. tuberculosis* infection or susceptibility to such an infection, "detecting an immune response" means examining the subject for signs of a immunological reaction to the administered material, e.g., reddening or swelling of the skin at the site of an intradermal injection. Where the subject has antibodies to the administered material, the

response will generally be rapid, e.g., 1 minute to 24 hours. On the other hand, a memory or activated T cell reaction of pre-immunized T lymphocytes in the subject is generally slower, appearing only after 24 hours and being maximal at 24-96 hours.

As used herein, a "subject" can be a human subject or a non-human mammal such as a non-human primate, a horse, a bovine animal, a pig, a sheep, a goat, a dog, a cat, a rabbit, a guinea pig, a hamster, a rat, or a mouse.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains. In case of conflict, the present document, including definitions, will control. Preferred methods and materials are described below, although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention. Unless otherwise indicated, these materials and methods are illustrative only and are not intended to be limiting. All publications, patent applications, patents and other references mentioned herein are illustrative only and not intended to be limiting.

Other features and advantages of the invention, e.g., methods of diagnosing or vaccinating against *M. tuberculosis* infection, will be apparent from the following description, from the drawings and from the claims.

## Brief Description of the Drawings

Figure 1 is a depiction of the amino acid sequences of *M. tuberculosis* polypeptides MTSP1-MTSP47.

Figure 2 is a depiction of the nucleotide sequences of the coding regions (mtsp1-mtsp47) encoding MTSP1-MTSP47.

Fig. 3A is a line graph showing the distribution of SPSCAN scores for the 3924 *M. tuberculosis* protein sequences obtained from the Sanger Center website.

Fig. 3B is a line graph showing the distribution of SignalP scores for the 3924 protein sequences obtained from the Sanger Center website.

Fig. 3C is a "dot plot" of SignalP scores versus SPSCAN scores for the individual 3924 protein sequences obtained from the Sanger Centre website.

Fig. 4 is an enlargement of Fig. 3C.

## Detailed Description

It is generally believed that proteins that are actively secreted by bacteria, especially intracellular bacteria (e.g., *Salmonella typhi* and *M. tuberculosis*), are effective as antigens that are capable of inducing protective immunity to the organism. A number of open reading frames (ORF), (i.e., DNA sequences that encode a protein) were predicted from the genomic sequence of *M. tuberculosis* [Cole et al. (1998) Nature 393:537-544]. The instant invention is based on the identification of a number of ORFs of this group that encode secreted polypeptides (see Example 1). The polypeptides encoded by the ORFs thus identified are designated *M. tuberculosis* Secreted Polypeptides (MTSP) and the DNA sequences encoding them are designated mtsp. Because they are secreted, we believe that the MTSP are both immunogenic and antigenic. The immune responses that they induce in subjects exposed to them are preferably also protective against *M. tuberculosis* infection in the subjects. The amino acid sequences of MTSP1-MTSP44 are shown in Fig. 1 and the nucleotide sequences of mtsp1-mtsp44 are shown in Fig. 2.

The invention encompasses: (a) isolated DNA molecules containing sequences (e.g., mtsp1-mtsp47) encoding polypeptides (e.g., MTSP1-MTSP47) secreted by *M. tuberculosis* and isolated portions of such DNA molecules that encode polypeptide segments having antigenic and immunogenic properties (i.e., functional segments); (b) the secreted polypeptides themselves (e.g., MTSP1-MTSP47) and functional segments of them; (c) antibodies (including antigen binding fragments, e.g., $F(ab')_2$, Fab, Fv, and single chain Fv fragments of such antibodies) that bind to the MTSP1-MTSP47 polypeptides and functional segments; (d) nucleic acid molecules (e.g., vectors) containing and capable of expressing one or more of the DNA molecules containing the mtsp1-mtsp47 sequences and portions of DNA molecules; (e) cells (e.g., bacterial, yeast, insect, or mammalian cells) transformed by such vectors; (f) compositions containing vectors encoding one or more *M. tuberculosis* polypeptides (or functional segments) including both the MTSP1-MTSP47 polypeptides (or functional segments thereof) and previously described *M. tuberculosis* polypeptides such as ESAT-6, 14 kDa antigen, MPT63, 19 kDa antigen, MPT64, MPT51, MTC28, 38 kDa antigen, 45/47 kDa antigen, MPB70, Ag85 complex, MPT53, and KatG (see also U.S. application no. 08/796,792); (g) compositions containing one or more *M. tuberculosis* polypeptides (or functional segments), including both the polypeptides of the invention and previously described *M. tuberculosis* polypeptides such as those described above; (h) compositions containing one or more of antibodies described in (c); (i) methods of diagnosis involving either (1) administration (e.g., intradermal injection) of the MTSP1-MTSP44 polypeptides of the invention, functional segments thereof, or mixtures of one more such polypeptides and/or

functional segments to a subject suspected of having or being susceptible to *M. tuberculosis* infection, (2) in vitro testing of lymphocytes from such a subject for responsiveness to the MTSP1-MTSP47 polypeptides, functional segments thereof, or the above mixtures, or (3) testing of a bodily fluid (e.g., blood, saliva, plasma, serum, urine, or semen or a lavage such as a bronchoalveolar lavage, a vaginal lavage, or lower gastrointestinal lavage) for antibodies to the MTSP1-MTSP47 polypeptides or functional segments thereof, or the above-described mixtures; (j) methods of vaccination involving administration to a subject of the compositions of either (f), (g), (h) or a combination of any two or even all 3 compositions.

With respect to diagnosis, purified *M. tuberculosis* proteins, functional segments of such proteins, or mixtures of proteins and/or the functional fragments have the advantage of discriminating infection by *M. tuberculosis* from infection by other bacteria, and in particular, non-pathogenic mycobacteria. Of particular benefit in such assays are proteins encoded by genes present in *M. tuberculosis*, and possibly other members of the M.tuberculosis complex (e.g., *M. tuberculosis*, M. bovis, M. microti, and M. africanum), but absent from the Bacille Calmette-Guerin (BCG) attenuated strain of M. bovis which has been commonly used for vaccination. Use of such proteins (e.g., the MTSP16 protein whose sequence is shown in Fig. 1) for diagnosis allows for discrimination between infection by *M. tuberculosis* and vaccination with BCG. Furthermore, compositions containing the *M. tuberculosis* proteins, functional segments of them, or mixtures of the proteins and/or the functional segments allows for improved quality control since "batch-to-batch" variability is greatly reduced

in comparison to complex mixtures such as purified protein derivative (PPD) of tuberculin.

Where vaccination is performed with nucleic acids both in vivo and ex vivo methods can be used. In vivo methods involve administration of the nucleic acids themselves to the subject and ex vivo methods involve obtaining cells (e.g., bone marrow cells or fibroblasts) from the subject, transducing the cells with the nucleic acids, preferably selecting or enriching for successfully transduced cells, and administering the transduced cells to the subject. Alternatively, the cells that are transduced and administered to the subject can be derived from another subject. Methods of vaccination and diagnosis are described in greater detail in U.S. application no. 08/796,792 which is incorporated herein by reference in its entirety.

The following example is meant to illustrate, not limit the invention.

**Example 1. Computer Aided Identification of *M. tuberculo sis* Secreted Proteins**

Software.

The software used to manipulate and analyze protein sequences was available from public web servers or was part of the Genetics Computer Group (GCG) package [Wisconsin Package Version 9.1, Genetics Computer Group (GCG), Madison, Wisc.]. Customized C-Shell scripts were used to automate some of the tasks or to extract selected information from the output of some of the programs. Signal peptides were predicted with SPSCAN, which is part of the GCG package, and SignalP, a program originating from the Center for Biological Sequence Analysis at the Technical University of Denmark, Lyngby, Denmark and currently available on the Internet at http://www.cbs.dtu.dk/services/SignalP. Putative

transmembrane segments were identified with the program TMpred and prokaryotic membrane lipoprotein lipid attachment sites with the program PrositeScan, both programs originating from the Bioinformatics Group at the Swiss Institute for Experimental Cancer Research in Epalinges, Switzerland, and currently available on the Internet at http://www.isrec.isb-sib.ch/software/TMPRED_form.html and http://www.isrec.isb-sib.ch/software/PSTSCAN_form.html, respectively. Protein similarity and relatedness was established with GAP and PILEUP, both in the GCG package, Blast originating from the National Center for Biotechnology Information of the National Institutes for Health, Bethesda, MD and currently available on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/, and AllAll originating from the Swiss Institute of Technology, Zurich, Switzerland, and currently available on the Internet at http:cbrg.inf.ethz.ch/subsection3_1_1.html.

Prediction of *M. tuberculosis* proteins with signal peptides

The amino acid sequences of the 3924 proteins predicted by the analysis of the *M. tuberculosis* genomic sequence have been made available by the Sanger Centre, Cambridge, England, and were downloaded from the current Sanger Center website [http://www.sanger.ac.uk/Projects/M_tuberculosis/]. Segments containing the first 70 amino acids of each predicted protein were analyzed by a system of our own design utilizing two different computer programs (SPSCAN and SignalP) designed to predict the occurrence of signal peptides. We concluded that combining the output from the two programs would increase the reliability of the selection. Both programs can detect signal peptides in polypeptides from eukaryotic and prokaryotic organisms, including gram-positive and gram-negative bacteria. To analyze the *M. tuberculosis* proteins the gram-positive mode was used. We performed an analysis

with SPSCAN allowing only one prediction per protein, setting the minimum score threshold at -10, both in the standard and the adjusted modes. In the adjusted mode, signal peptides longer than a certain threshold value are penalized. We found that the correlation between the scores obtained with SPSCAN in the standard and adjusted modes increased with the value of the score, i.e., signal peptides that received high scores in standard mode also had high scores in the adjusted mode. We determined to use only the adjusted mode in subsequent steps.

To define cutoff values for the scores obtained with SPSCAN (in adjusted mode) and SignalP we took into account the following factors: (a) SignalP scores above 0.34 are generally considered significant; (b) the analysis of Haemophilus influenzae genome with SignalP yielded the prediction that about 10% of the encoded proteins contain a signal peptide; and (c) the average scores of thirteen known secreted or membrane-associated *M. tuberculosis* antigens was 9.11 (standard deviation (SD)=1.8) and 0.55 (SD=0.15), as calculated as above utilizing SPSCAN and SignalP, respectively (Table 1).

Of the 3924 *M. tuberculosis* protein sequences downloaded from the Sanger Centre website, about 10% of the sequences had SPSCAN scores equal or higher than 8 (Fig. 3A) and about 10% of the sequences had SignalP scores equal or higher than 0.4 (Fig 3B). We tentatively adopted these score values as "cutoffs" and we used the cutoffs to construct a list of proteins that were likely to be either secreted or exposed at the bacterial cell surface. This list included those proteins with SPSCAN scores higher than 8 and SignalP scores higher than 0.4. We refer to this group of proteins (208

entries, about 5% of the proteome) as the "Top208" group (Fig. 3C and Fig. 4).

Table 1. SPCAN and SignalP Scores of Known Secreted or Membrane Associated *M. tuberculosis* Polypeptide Antigens

| Polypeptide Antigens | Alternative Names | SPSCAN Score | SignalP Score |
|---|---|---|---|
| 19 kDa | | 5.9 | 0.331 |
| 38 kDa | PhoS, Ag78, antigen 5 | 6.3 | 0.505 |
| 45/47 kDa | | 11.2 | 0.627 |
| MPT44 | Ag85A, P32, FbpA | 9.2 | 0.425 |
| MPT45 | Ag85C, FbpC | 10.1 | 0.496 |
| MPT51 | | 11 | 0.758 |
| MPT53 | | 9.4 | 0.581 |
| MPT59 | Ag85B, á antigen, Ag 6, FbpB | 9.7 | 0.629 |
| MPT63 | | 8 | 0.57 |
| MPT64 | | 10.2 | 0.83 |
| MPT70 | | 9 | 0.459 |
| MPT83 | | 7.1 | 0.298 |
| MTC28 | | 11.4 | 0.7 |

Prediction of *M. tuberculosis* secreted proteins

A signal peptide may target a protein to the membrane but does not define a secreted protein, because additional transmembrane segments within the mature protein molecule can be present. In addition, lipoproteins are also targeted to the membrane by a signal peptide, but are not all secreted since cleavage of the signal peptide is coupled with the attachment of an acyl glycerol group that anchors the protein to the membrane. In light of these considerations and the fact that SignalP is not designed to differentiate lipoprotein signal peptides from secretory signal peptides, we believe that the Top208 group contains lipoproteins and proteins with multiple transmembrane segments, in addition to secreted proteins.

The number of putative transmembrane segments and the presence of lipoprotein lipid attachment sites were assessed by analyzing the Top208 proteins with TMpred and PrositeScan.

TMpred identifies putative transmembrane segments by comparing a query amino acid sequence with a database of amino acid sequences of experimentally defined transmembrane segments. Scores higher than 500 are considered significant. PrositeScan compares query amino acid sequences against the Prosite database of protein motifs. The prokaryotic lipoprotein lipid attachment site motif is entry number PS00013. Our methodology identified a class of secreted proteins (the "Top208-TM1" group that included MTSP1-MTSP44) which were characterized by a single transmembrane segment (with score higher than 500) in the position predicted for the signal peptide and in which no lipoprotein motifs were identified. Other proteins had additional transmembrane segments with scores higher than 500, had lipoprotein motifs, or were excluded from the analysis because they belonged to the PE/PPE/PGRS families of proteins [Cole et al., 1998] and their biased amino acid composition made it difficult to obtain reliable results with SPSCAN, SignalP, or TMpred. A summary of the characteristics of the proteins we assigned to the Top208-TM1 group is presented in Table 2 and data regarding proteins MTSP1-MTSP47 are presented in Table 3. The amino acid sequences of the proteins are listed in Fig. 1 and the nucleotide sequences of ORF encoding them (mtsp1-mtsp47) are listed in Fig. 2.

Table 2. Features defining the *M. tuberculosis* proteins included in the Top208-TM1 group.

1. A signal peptide with score higher than 0.4 was predicted with SignalP in the first 70 amino acids.
2. A signal peptide with score higher than 8 was predicted with SPSCAN in the first 70 amino acids.
3. A single transmembrane segment, with a score greater than 500 and coinciding approximately with the putative signal peptide, was predicted by TMpred.
4. No lipoprotein lipid attachment sites were identified with PrositeScan.

Table 3. Proteins included in the Top208-TM1 group.

| Protein | No. of Amino Acids | SPSCAN Score | SPSCAN Sequence | SignalP Score | SignalP Sequence |
|---|---|---|---|---|---|
| MTSP20 | 130 | 12.4 | 1-32 | 0.672 | 1-32 |
| MTSP21 | 109 | 8.4 | 1-22 | 0.631 | 1-22 |
| MTSP23 | 114 | 10.2 | 1-34 | 0.592 | 1-34 |
| MTSP16 | 126 | 9.2 | 1-28 | 0.557 | 1-36 |
| MTSP24 | 125 | 11.4 | 1-35 | 0.73 | 1-35 |
| MTSP14 | 144 | 8.9 | 1-34 | 0.584 | 1-34 |
| MTSP13 | 157 | 10 | 1-32 | 0.753 | 1-32 |
| MTSP22 | 124 | 8.6 | 1-30 | 0.592 | 1-30 |
| MTSP25 | 155 | 9.5 | 35-49 | 0.842 | 1-49 |
| MTSP27 | 233 | 13.8 | 1-29 | 0.787 | 1-29 |
| MTSP11 | 233 | 10.9 | 1-32 | 0.779 | 1-32 |
| MTSP26 | 382 | 8.3 | 1-34 | 0.721 | 1-34 |
| MTSP12 | 214 | 12.6 | 1-28 | 0.71 | 1-28 |
| MTSP8 | 158 | 9.1 | 1-33 | 0.695 | 1-30 |
| MTSP10 | 155 | 8.8 | 15-45 | 0.669 | 1-45 |
| MTSP28 | 295 | 14.8 | 1-31 | 0.667 | 1-31 |
| MTSP9 | 241 | 10 | 1-22 | 0.635 | 1-22 |
| MTSP29 | 380 | 12.4 | 1-27 | 0.621 | 1-27 |
| MTSP2 | 111 | 10.6 | 1-28 | 0.579 | 1-28 |
| MTSP4 | 177 | 8.7 | 1-25 | 0.578 | 1-24 |
| MTSP17 | 219 | 8.9 | 1-29 | 0.543 | 1-29 |
| MTSP3 | 282 | 11.5 | 1-32 | 0.538 | 1-32 |
| MTSP18 | 220 | 8.8 | 38-68 | 0.537 | 1-68 |
| MTSP6 | 219 | 8.4 | 1-34 | 0.537 | 1-34 |
| MTSP7 | 136 | 11.7 | 1-24 | 0.53 | 1-24 |
| MTSP31 | 457 | 9.1 | 1-18 | 0.494 | 1-25 |
| MTSP30 | 286 | 8.3 | 15-37 | 0.469 | 1-37 |
| MTSP1 | 104 | 8.2 | 1-28 | 0.466 | 1-28 |
| MTSP15 | 134 | 10 | 1-21 | 0.458 | 1-56 |
| MTSP32 | 449 | 8.8 | 1-23 | 0.444 | 1-23 |
| MTSP19 | 169 | 10.5 | 28-53 | 0.438 | 1-53 |
| MTSP5 | 568 | 9.9 | 1-31 | 0.432 | 1-31 |
| MTSP33 | 113 | 11.9 | 1-25 | 0.873 | 1-25 |
| MTSP41 | 112 | 12 | 1-33 | 0.663 | 1-3 |
| MTSP38 | 173 | 10.5 | 1-28 | 0.697 | 1-28 |
| MTSP35 | 408 | 8.8 | 1-33 | 0.616 | 1-33 |
| MTSP34 | 149 | 13.7 | 1-23 | 0.888 | 1-23 |
| MTSP36 | 168 | 11.3 | 1-28 | 0.824 | 1-27 |
| MTSP42 | 521 | 8.4 | 1-34 | 0.679 | 1-34 |
| MTSP44 | 149 | 11 | 1-30 | 0.661 | 1-30 |
| MTSP37 | 228 | 9.4 | 1-23 | 0.598 | 1-23 |
| MTSP40 | 231 | 9.2 | 1-30 | 0.55 | 1-30 |
| MTSP43 | 137 | 8.2 | 1-36 | 0.485 | 1-37 |
| MTSP39 | 509 | 8.6 | 1-35 | 0.413 | 1-38 |
| MTSP45 | 145 | 8.4 | 1-46 | 0.412 | 1-62 |
| MTSP46 | 143 | 8.5 | 1-27 | 0.555 | 1-66 |
| MTSP47 | 171 | 8.3 | 1-35 | 0.424 | 1-30 |

SPSCAN sequence and SignalP sequence show the sequence, in terms of amino acid residue numbers, included in the signal peptide predicted by SPSCAN and SignalP, respectively.

Table 4. Presence mtsp coding regions in various strains of *Mycobacterium tuberculosis*.

| Coding Region | *M. tuberculosis* | *M. bovis BCG* | *M. bovis* | *M. kansasii* | *M. africanum* | *M. scrofulaceum* | *M. fortuitum* | *M. marinum* | *M. malmoense* | *M. avium* | *M. gastri* | *M. chelonae* | *M. ulcerans* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MTSP6 | + | + | + | + | + | + | + | + | + | + | + | + | + |
| MTSP28 | + | + | + | + | + | + | + | + | + | + | + | + | + |
| MTSP44 | + | + | + | + | + | + | + | + | + | + | + | + | +/- |
| MTSP34 | + | + | + | +/- | + | + | + | + | + | + | + | + | + |
| MTSP39 | + | + | + | + | + | + | + | + | + | + | + | + | + |
| MTSP1 | + | + | + | + | + | + | + | + | + | + | + | + | + |
| MTSP15 | + | + | + | + | + | + | + | + | + | + | + | + | + |
| MTSP35 | + | + | + | + | + | + | + | + | + | + | + | + | + |
| MTSP5 | + | + | + | + | + | + | + | + | + | + | + | + | + |
| MTSP46 | + | + | + | + | + | - | - | + | - | - | + | - | - |
| MTSP11 | + | + | + | + | + | - | - | - | + | - | + | - | - |
| MTSP24 | + | + | + | + | + | + | - | + | + | - | + | + | + |
| MTSP23 | + | + | + | + | + | + | - | + | + | + | + | - | - |
| MTSP41 | + | + | + | + | + | - | - | - | + | - | + | - | - |
| MTSP22 | + | + | + | - | + | - | + | - | - | - | - | - | - |
| MTSP26 | + | + | + | + | + | + | + | + | + | + | + | + | + |
| MTSP40 | + | + | + | + | + | + | + | + | + | + | + | + | + |
| MTSP13 | + | + | + | + | + | + | + | + | + | + | - | - | - |
| MTSP16 | + | - | - | - | + | - | - | - | - | - | - | - | - |
| MTSP42 | + | + | + | + | + | + | - | + | + | - | + | + | + |
| MTSP36 | + | + | + | + | + | | | | | | | | |
| MTSP47 | + | + | + | - | + | - | - | - | - | - | - | - | - |
| MTSP38 | + | + | + | + | + | + | + | + | + | + | + | + | + |
| MTSP10 | + | + | + | + | + | + | + | + | + | + | + | + | + |
| MTSP37 | + | + | + | + | + | | | | | | | | |
| MTSP29 | + | + | + | + | + | + | + | + | + | + | + | +/- | +/- |
| MTSP31 | + | + | + | + | + | + | + | + | + | + | + | + | + |
| MTSP32 | + | + | + | + | + | + | + | + | + | + | + | + | + |
| MTSP30 | + | + | + | + | + | + | + | + | + | + | +/- | +/- | + |
| MTSP3 | + | + | + | + | + | + | + | + | + | + | + | + | + |
| MTSP20 | + | + | + | - | + | | | | | | | | |
| MTSP4 | + | + | + | + | + | + | + | + | + | + | + | + | + |
| MTSP27 | + | + | + | + | + | + | + | + | + | + | + | +/- | +/- |

The inventors have found, by standard DNA hybridization. Southern blotting techniques using the indicated coding regions as probes and DNA isolated from the indicated strains of *Mycobacteria*, that some of the coding regions are specific for the *M. tuberculosis* complex. (Table 4)

Although the invention has been described with reference to the presently preferred embodiment, it should be understood that various modifications can be made without departing from the spirit of the invention. Accordingly, the invention is limited only by the following claims.

What is claimed is:

1. An isolated DNA molecule comprising a DNA sequence encoding a polypeptide with a first amino acid sequence selected from the group consisting of the amino acid sequences of the polypeptides MTSP1, MTSP2, MTSP3, MTSP4, MTSP5, MTSP6, MTSP7, MTSP8, MTSP9, MTSP10, MTSP11, MTSP12, MTSP13, MTSP14, MTSP15, MTSP16, MTSP17, MTSP18, MTSP19, MTSP20, MTSP21, MTSP22, MTSP23, MTSP24, MTSP25, MTSP26, MTSP27, MTSP28, MTSP29, MTSP30, MTSP31, MTSP32, MTSP33, MTSP34, MTSP35, MTSP36, MTSP37, MTSP38, MTSP39, MTSP40, MTSP41, MTSP42, MTSP43, MTSP44, MTSP45, MTSP46, and MTSP47 as depicted in Fig. 1,

or a second amino acid sequence identical to said first amino acid sequence but with conservative substitutions,

wherein said polypeptide has *Mycobacterium tuberculosis* specific antigenic and immunogenic properties.

2. An isolated portion of the DNA molecule of claim 1, said portion encoding a segment of said polypeptide shorter than the full-length polypeptide, said segment having *Mycobacterium tuberculosis* specific antigenic and immunogenic properties.

3. A vector comprising:

(a) the DNA molecule of claim 1; and

(b) transcriptional and translational regulatory sequences operationally linked to said DNA sequence, said regulatory sequences allowing for expression of the polypeptide encoded by said DNA sequence in a cell.

4. A vector comprising:

(a) the DNA molecule of claim 2; and

(b) transcriptional and translational regulatory sequences operationally linked to said DNA sequence, said regulatory sequences allowing for expression of the polypeptide encoded by said DNA sequence in a cell.

5. A cell transformed with the vector of claim 3.

6. A cell transformed with the vector of claim 4.

7. A composition comprising the vector of claim 3 and a pharmaceutically acceptable diluent or filler.

8. A composition comprising the vector of claim 4 and a pharmaceutically acceptable diluent or filler.

9. A composition for use as a DNA vaccine, said composition comprising at least two DNA sequences, each encoding a polypeptide of the *Mycobacterium tuberculosis* complex or a functional segment thereof, said DNA sequences being operationally linked to transcriptional and translational regulatory sequences which allow for expression of each said polypeptide in a cell of a vertebrate,

wherein at least one of said DNA sequences is the sequence of claim 1.

10. A composition for use as a DNA vaccine, said composition comprising at least two DNA sequences, each encoding a polypeptide of the *Mycobacterium tuberculosis* complex or a functional segment thereof, said DNA sequences being operationally linked to transcriptional and translational regulatory sequences which allow for expression of each said polypeptide in a cell of a vertebrate,

wherein at least one of said DNA sequences is the sequence of claim 2.

11. An isolated polypeptide with a first amino acid sequence selected from the group consisting of the sequences of the polypeptides MTSP1, MTSP2, MTSP3, MTSP4, MTSP5, MTSP6, MTSP7, MTSP8, MTSP9, MTSP10, MTSP11, MTSP12, MTSP13, MTSP14, MTSP15, MTSP16, MTSP17, MTSP18, MTSP19, MTSP20, MTSP21, MTSP22, MTSP23, MTSP24, MTSP25, MTSP26, MTSP27, MTSP28, MTSP29, MTSP30, MTSP31, MTSP32, MTSP33, MTSP34, MTSP35, MTSP36, MTSP37, MTSP38, MTSP39, MTSP40, MTSP41, MTSP42, MTSP43, MTSP44, mtsp45, mtsp46, and MTSP47 as depicted in Fig. 1,

or a second amino acid sequence identical to said first amino acid sequence but with conservative substitutions,

wherein said polypeptide has *Mycobacterium tuberculosis* specific antigenic and immunogenic properties.

12. An isolated segment of the polypeptide of claim 11, said segment being shorter than the full-length polypeptide and having *Mycobacterium tuberculosis* specific antigenic and immunogenic properties.

13. A composition comprising the polypeptide of claim 11, or a functional segment thereof, and a pharmaceutically acceptable diluent or filler.

14. A composition comprising the polypeptide of claim 12, or a functional segment thereof, and a pharmaceutically acceptable diluent or filler.

15. A composition comprising at least two polypeptides of the *Mycobacterium tuberculosis* complex, or functional segments thereof, wherein at least one of said at least two polypeptides is the sequence of claim 1.

16. A composition comprising at least two polypeptides of the *Mycobacterium tuberculosis* complex, or functional segments thereof, wherein at least one of said at least polypeptides is the segment of claim 2.

17. A method of diagnosis comprising:

(a) administration of the composition of claim 13 to a subject suspected of having or being susceptible to *Mycobacterium tuberculosis* infection; and

(b) detecting an immune response in said subject to said composition, as an indication that said subject has or is susceptible to *Mycobacterium tuberculosis* infection.

18. A method of diagnosis comprising:

(a) administration of the composition of claim 14 to a subject suspected of having or being susceptible to *Mycobacterium tuberculosis* infection; and

(b) detecting an immune response in said subject to said composition, as an indication that said subject has or is susceptible to *Mycobacterium tuberculosis* infection.

19. A method of diagnosis comprising:

(a) administration of the composition of claim 15 to a subject suspected of having or being susceptible to *Mycobacterium tuberculosis* infection; and

(b) detecting an immune response in said subject to said composition as an indication that said subject has or is susceptible to *Mycobacterium tuberculosis* infection.

20. A method of diagnosis comprising:

(a) administration of the composition of claim 16 to a subject suspected of having or being susceptible to *Mycobacterium tuberculosis* infection; and

(b) detecting an immune response in said subject to said composition as an indication that said subject has or is susceptible to *Mycobacterium tuberculosis* infection.

21. A method of diagnosis comprising:

(a) providing a population of cells comprising CD4 T lymphocytes from a subject;

(b) providing a population of cells comprising antigen presenting cells (APC) expressing a major histocompatibility complex (MHC) class II molecule expressed by said subject;

(c) contacting the CD4 lymphocytes of (a) with the APC of (b) in the presence of the polypeptide of claim 1; and

(d) determining the ability of said CD4 lymphocytes to respond to said polypeptide, as an indication that said subject has or is susceptible to *Mycobacterium tuberculosis* infection.

22. A method of diagnosis comprising:

(a) providing a population of cells comprising CD4 T lymphocytes from a subject;

(b) providing a population of cells comprising antigen presenting cells (APC) expressing at least one major histocompatibility complex (MHC) class II molecule expressed by said subject;

(c) contacting the CD4 lymphocytes of (a) with the APC of (b) in the presence of the segment of claim 2; and

(d) determining the ability of said CD4 lymphocytes to respond to said polypeptide, as an indication that said subject has or is susceptible to *Mycobacterium tuberculosis* infection.

23. A method of diagnosis comprising:

(a) providing a population of cells comprising CD4 T lymphocytes from a subject;

(b) providing a population of cells comprising antigen presenting cells (APC) expressing at least one major histocompatibility complex (MHC) class II molecule expressed by said subject;

(c) contacting the CD4 lymphocytes of (a) with the APC of (b) in the presence of the composition of claim 15; and

(d) determining the ability of said CD4 lymphocytes to respond to said polypeptide, as an indication that said subject has or is susceptible to *Mycobacterium tuberculosis* infection.

24. A method of diagnosis comprising:

(a) providing a population of cells comprising CD4 T lymphocytes from a subject;

(b) providing a population of cells comprising antigen presenting cells (APC) expressing at least one major histocompatibility complex (MHC) class II molecule expressed by said subject;

(c) contacting the CD4 lymphocytes of (a) with the APC of (b) in the presence of the composition of claim 16; and

(d) determining the ability of said CD4 lymphocytes to respond to said polypeptide, as an indication that said subject has or is susceptible to *Mycobacterium tuberculosis* infection.

25. A method of diagnosis comprising:

(a) contacting the polypeptide of claim 11 with a bodily fluid of a subject;

(b) detecting the presence of binding of antibody to said polypeptide, as an indication that said subject has or is susceptible to *Mycobacterium tuberculosis* infection.

26. A method of diagnosis comprising:

(a) contacting the segment of claim 12 with a bodily fluid of a subject;

(b) detecting the presence of binding of antibody to said polypeptide, as an indication that said subject has or is susceptible to *Mycobacterium tuberculosis* infection.

27. A method of diagnosis comprising:

(a) contacting the composition of claim 15 with a bodily fluid of a subject;

(b) detecting the presence of binding of antibody to said composition, as an indication that said subject has or is susceptible to *Mycobacterium tuberculosis* infection.

28. A method of diagnosis comprising:

(a) contacting the composition of claim 16 with a bodily fluid of a subject;

(b) detecting the presence of binding of antibody to said composition, as an indication that said subject has or is susceptible to *Mycobacterium tuberculosis* infection.

29. A method of vaccination comprising administration of the composition of claim 7 to a subject.

30. A method of vaccination comprising administration of the composition of claim 8 to a subject.

31. A method of vaccination comprising administration of the composition of claim 9 to a subject.

32. A method of vaccination comprising administration of the composition of claim 10 to a subject.

33. A method of vaccination comprising administration of the composition of claim 13 to a subject.

34. A method of vaccination comprising administration of the composition of claim 14 to a subject.

35. A method of vaccination comprising administration of the composition of claim 15 to a subject.

36. A method of vaccination comprising administration of the composition of claim 16 to a subject.

FIG. 1

MTSP1
MNRIVQFGVSAVAAAAIGIGAGSGIAAAFDGEDEVTGPDADRARAAAVQAVPGGTAGEVE
TETGEGAAAYGVLVTRPDGTRVEVHLDRDFRVLDTEPADGDGG*

MTSP2
MRLSLTALSAGVGAVAMSLTVGAGVASADPVDAVINTTCNYGQVVAALNATDPGAAAQFN
ASPVAQSYLRNFLAAPPPQRAAMAAQLQAVPGAAQYIGLVESVAGSCNNY*

MTSP3
MFTGIASHAGALGAALVVLIGAAILHDGPAAADPNQDDRFLALLEKKEIPAVANVPRVID
AAHKVCRKLDGGMPVNDIVDGLRNDAYNIDPVMRLYPVRLTTTMTRFISAAVEIYCPNHH
SKMAFAMANFEPGSNEPTHRVAASTRSAVNSGSDLRASVSDMTIMSPGWREPTGAMLASV
LGAVRAGDPLIPNPPPIPVPPPAAQTLIPPPPIVAPPPPRPAPPQQPPPPPPEVEPPAGV
PQSGGAAGSGGAGSGGGGGGDGPVEPSPARPMPPGFIRLAP*

MTSP4
MTRLIPGCTLVGLMLTLLPAPTSAAGSNTATTLFPVDEVTQLETHTFLDCHPNGSCDFVA
GANLRTPDGPTGFPPGLWARQTTEIRSTNRLAYLDAHATSQFERVMKAGGSDVITTVYFG
EGPPDKYQTTGVIDSTNWSTGQPMTDVNVIVCTHMQVVYPGVNLTSPSTCAQANFS*

MTSP5
MVLRSRKSTLGVVVCLALVLGGPLNGCSSSASHRGPLNAMGSPAIPSTAQEIPNPLRGQY
EDLMEPLFPQGNPAQQRYPPWPASYDASLRVSWRQLQPTDPRTLPPDAPDDRKYDFSVID
NALTRLADRGMRLTLRVYAYSSCCKASYPDGTNIAIPDWERAIASTNTSYPGPATDPSTG
VVQVVPNFNDSTYLNDFAQLLAALGRRYDGDERLSVFEFSGYGDFSENHVAYLRDTLGAP
GPGPDESVATLGYYSQFRDQNITTASIKQLIAANVSAFPHTQLVTSPANPEIVRELFADE
VTNKLAAPVGVRSDCLGVDAPLPAWAESSTSHYVQTKDPVVAALRQRLATAPVITEWCEL
PTGSSPRAYYEKGLRDVIRYHVSMTSSVNFPDQTATSPMDPALYLVWAQANAAAGYRYSV
EAQPGSQALAGKVATISVTWTNYGAAAATEKWVPGYRLVDSTGQVVRTLPAAVDLKTLVS
DQRGDRSSDQPTPASVAETVRVDLSGLPAGHYTLRAAIDWQQHKPNGSHVVNYPPMLLSR
DGRDDSGFYPVATLDIPRDAQTAVNAS*

MTSP6
MSRLLALLCAAVCTGCVAVVLAPVSLAVVNPWFANSVGNATQVVSVVGTGGSTAKMDVYQ
RTAAGWQPLKTGITTHIGSAGMAPEAKSGYPATPMGVYSLDSAFGTAPNPGGGLPYTQVG
PNHWWSGDDNSPTFNSMQVCQKSQCPFSTADSENLQIPQYKHSVVMGVNKAKVPGKGSAF
FFHTTDGGPTAGCVAIDDATLVQIIRWLRPGAVIAIAK*

MTSP7
MIRELVTTAAITGAAIGGAPVAGADPQRYDGDVPGMNYDASLGAPCSSWERFIFGRGPSG
QAEACHFPPPNQFPPAETGYWVISYPLYGVQQVGAPCPKPQAAAQSPDGLPMLCLGARGW
QPGWFTGAGFFPPEP*

FIG. 1 (continued)

MTSP8
MGELRLVGGVLRVLVVVGAVFDVAVLNAGAASADGPVQLKSRLGDVCLDAPSGSWFSPLV
INPCNGTDFQRWNLTDDRQVESVAFPGECVNIGNALWARLQPCVNWISQHWTVQPDGLVK
SDLDACLTVLGGPDPGTWVSTRWCDPNAPDQQWDSVP*

MTSP9
MPAMTARSVVLSVLLGAHPAWATASELIQLTADFGIKETTLRVALTRMVGAGDLVRSADG
YRLSDRLLARQRRQDEAMRPRTRAWHGNWHMLIVTSIGTDARTRAALRTCMHHKRFGELR
EGVWMRPDNLDLDLESDVAARVRMLTARDEAPADLAGQLWDLSGWTEAGHRLLGDMAAAT
DMPGRFVVAAAMVRHLLTDPMLPAELLPADWPGAGLRAAYHDFATAMAKRRDATQLLEVT
*

MTSP10
VPAGVGNASGSVLDMTSVRTVPSAVALVTFAGAALSGVIPAIARADPVGHQVTYTVTTTS
DLMANIRYMSADPPSMAAFNADSSKYMITLHTPIAGGQPLVYTATLANPSQWAIVTASGG
LRVNPEFHCEIVVDGQVVVSQDGGSGVQCSTRPW*

MTSP11
MTTSKIATAFKTATFALAAGAVALGLASPADAAAGTMYGDPAAAAKYWRQQTYDDCVLMS
AADVIGQVTGREPSERAIIKVAQSTPSVVHPGSIYTKPADAEHPNSGMGTSVADIPTLLA
HYGVDAVITDEDHATATGVATGMAALEQYLGSGHAVIVSINAEMIWGQPVEETDSAGNPR
SDHAVVVTGVDTENGIVHLNDSGTPTGRDEQIPMETFVEAWATSHDFMAVTT*

MTSP12
MGVIARVVGVAACGLSLAVLAAAPTAGAEPTGALPPMTSSGSGPVIGDGDAALRQRISQQ
LFSFGDPTVQEVDGSDAAQFITAAAAVADRDVASVFLPLQRVLGCQQNTAGSGAGFGARA
YRRTDGQWGGAMLVVAKSTVSDVDALKACVKSGWRKATAGTPTSMCNNGWTYPPFADTRR
GEEGYFVLLAGTASDFCSAPNANYRTTASSWPG*

MTSP13
MRLKPAPSPAAAFAVAGLILAGWAGSVGLAGADPEPAPTPKTAIDSDGTYAVGIDIAPGT
YSSAGPVGDGTCYWKRMGNPDGALIDNALSKKPQVVTIEPTDKAFKTHGCQPWQNTGSEG
AAPAGVPGPEAGAQLQNQLGILNGLLGPTGGRVPQP*

MTSP14
MITNLRRRTAMAAAGLGAALGLGILLVPTVDAHLANGSMSEVMMSEIAGLPIPPIIHYGA
IAYAPSGASGKAWHQRTPARAEQVALEKCGDKTCKVVSRFTRCGAVAYNGSKYQGGTGLT
RRAAEDDAVNRLEGGRIVNWACN*

MTSP15
VTVLLDANVLIALVVAEHVHHDAAADWLMASDTGFATCPMTQGSLVRFLVRSGQSAAAAR
DVVSAVQCTSRHEFWPDALSFAGVEVAGVVGHRQVTDAYLAQLARSHDGQLATLDSGLAH
LHGDVAVLIPTTT*

FIG. 1 (continued)

MTSP16
VQRQSLMPQQTLAAGVFVGALLCGVVTAAVPPHARADVVAYLVNVTVRPGYNFANADAAL
SYGHGLCEKVSRGRPYAQIIADVKADFDTRDQYQASYLLSQAVNELCPALIWQLRNSAVD
NRRSG*

MTSP17
VRSYLLRIELADRPGSLGSLAVALGSVGADILSLDVVERGNGYAIDDLVVELPPGAMPDT
LITAAEALNGVRVDSVRPHTGLLEAHRELELLDHVAAAEGATARLQVLVNEAPRVLRVSW
CTVLRSSGGELHRLAGSPGAPETRANSAPWLPIERAAALDGGADWVPQAWRDMDTTMVAA
PLGDTHTAVVLGRPGPEFRPSEVARLGYLAGIVATMLR*

MTSP18
MPDGEQSQPPAQEDAEDDSRPDAAEAAAAEPKSSAGPMFSTYGIASTLLGVLSVAAVVLG
AMIWSAHRDDSGERTYLTRVMLTAAEWTAVLINMNADNIDASLQRLHDGTVGQLNTDFDA
VVQPYRQVVEKLRTHSSGRIEAVAIDTVHRELDTQSGAARPVVTTKLPPFATRTDSVLLV
ATSVSENAGAKPQTVHWNLRLDVSDVDGKLMISRLESIR*

MTSP19
MKMVKSIAAGLTAAAAIGAAAAGVTSIMAGGPVVYQMQPVVFGAPLPLDPASAPDVPTAA
QLTSLLNSLADPNVSFANKGSLVEGGIGGTEARIADHKLKKAAEHGDLPLSFSVTNIQPA
AAGSATADVSVSGPKLSSPVTQNVTFVNQGGWMLSRASAMELLQAAGN*

MTSP20
MNLRRHQTLTLRLLAASAGILSAAAFAAPAQANPVDDAFIAALNNAGVNYGDPVDAKALG
QSVCPILAEPGGSFNTAVASVVARAQGMSQDMAQTFTSIAISMYCPSVMADVASGNLPAL
PDMPGLPGS*

MTSP21
MRVVSTLLSIPLMIGLAVPAHAGPSGDDAVFLASLERAGITYSHPDQAIASGKAVCALVE
SGESGLQVVNELRTRNPGFSMDGCCKFAAISAHVYCPHQITKTSVSAK*

MTSP22
MARTLALRASAGLVAGMAMAAITLAPGARAETGEQFPGDGVFLVGTDIAPGTYRTEGPSN
PLILVFGRVSELSTCSWSTHSAPEVSNENIVDTNTSMGPMSVVIPPTVAAFQTHNCKLWM
RIS*

MTSP23
MLSPLSPRIIAAFTTAVGAAAIGLAVATAGTAGANTKDEAFIAQMESIGVTFSSPQVATQ
QAQLVCKKLASGETGTEIAEEVLSQTNLTTKQAAYFVVDATKAYCPQYASQLT*

FIG. 1 (continued)

MTSP24
MTTMITLRRRFAVAVAGVATAAATTVTLAPAPANAADVYGAIAYSGNGSWGRSWDYPTRA
AAEATAVKSCGYSDCKVLTSFTACGAVAANDRAYQGGVGPTLAAAMKDALTKLGGGYIDT
WACN*

MTSP25
MTPGLLTTAGAGRPRDRCARIVCTVFIETAVVATMFVALLGLSTISSKADDIDWDAIAQC
ESGGNWAANTGNGLYGGLQISQATWDSNGGVGSPAAASPQQQIEVADNIMKTQGPGAWPK
CSSCSQGDAPLGSLTHILTFLAAETGGCSGSRDD*

MTSP26
VQGAVAGLVFLAVLVIFAIIVVAKSVALIPQAEAAVIERLGRYSRTVSGQLTLLVPFIDR
VRARVDLRERVVSFPPQPVITEDNLTLNIDTVVYFQVTVPQAAVYEISNYIVGVEQLTTT
TLRNVVGGMTLEQTLTSRDQINAQLRGVLDEATGRWGLRVARVELRSIDPPPSIQASMEK
QMKADREKRAMILTAEGTREAAIKQAEGQKQAQILAAEGAKQAAILAAEADRQSRMLRAQ
GERAAAYLQAQGQAKAIEKTFAAIKAGRPTPEMLAYQYLQTLPEMARGDANKVWVVPSDF
NAALQGFTRLLGKPGEDGVFRFEPSPVEDQPKHAADGDDAEVAGWFSTDTDPSIARAVAT
AEAIARKPVEGSLGTPPRLTQ*

MTSP27
LQTAHRRFAAAFAAVLLAVVCLPANTAAADDKLPLGGGAGIVVNGDTMCTLTTIGHDKNG
DLIGFTSAHCGGPGAQIAAEGAENAGPVGIMVAGNDGLDYAVIKFDPAKVTPVAVFNGFA
INGIGPDPSFGQIACKQGRTTGNSCGVTWGPGESPGTLVMQVCGGPGDSGAPVTVDNLLV
GMIHGAFSDNLPSCITKYIPLHTPAVVMSINADLADINAKNRPGAGFVPVPA*

MTSP28
MLMPEMDRRRMMMMAGFGALAAALPAPTAWADPSRPAAPAGPTPAPAAPAAATGGLLFHD
EFDGPAGSVPDPSKWQVSNHRTPIKNPVGFDRPQFFGQYRDSRQNVFLDGNSNLVLRATR
EGNRYFGGLVHGLWRGGIGTTWEARIKFNCLAPGMWPAWWLSNDDPGRSGEIDLIEWYGN
GTWPSGTTVHANPDGTAFETCPIGVDGGWHNWRVTWNPSGMYFWLDYADGIEPYFSVPAT
GIEDLNEPIREWPFNDPGYKVFPVLNLAVGGSGGGDPATGSYPQEMLVDWVRVF*

MTSP29
VHRRTALKLPLLLAAGTVLGQAPRAAAEEPGRWSADRAHRWYQAHGWLVGANYITSNAIN
QLEMFQPGTYDPRRIDNELGLARFHGFNTVRVFLHDLLWAQDAPGFQTRLAQFVAIAARY
HIKPLFVLFDSCWDPLPRPGRQRAPRAGVHNSGWVQSPGAERLDDRRYASTLYNYVTGVL
GQFRNDDRVLGWDLWNEPDNPARVYRKVERKDKLERVAELLPQVFRWARTVDPVQPLTSG
VWQGNWGDPGRRSTISAIQLDNADVITFHSYAAPAEFEGRIAELAPLQRPILCTEYLARS
QGSTVEGILPIAKRHNVGAFNWGLVAGKTQTYLPWDSWDHPYRAPPKVWFHDLLHPNGRP
YRDGEVQTIRKLNGMPSQD*

MTSP30
VSTYGWRAYALPVLMVLTTVVVYQTVTGTSTPRPAAAQTVRDSPAIGVVGTAILDAPPRG
LAVFDANLPAGTLPDGGPFTEAGDKTWRVVPGTTPQVGQGTVKVFRYTVEIENGLDPTMY
GGDNAFAQMVDQTLTNPKGWTHNPQFAFVRIDSGKPDFRISLVSPTTVRGGCGYEFRLET
SCYNPSFGGMDRQSRVFINEARWVRGAVPFEGDVGSYRQYVINHEVGHAIGYLRHEPCDQ
QGGLAPVMMQQTFSTSNDDAAKFDPDFVKADGKTCRFNPWPYPIP*

FIG. 1 (continued)

MTSP31
MRPYYIAIVGSGPSAFFAAASLLKAADTTEDLDMAVDMLEMLPTPWGLVRSGVAPDHPKI
KSISKQFEKTAEDPRFRFFGNVVVGEHVQPGELSERYDAVIYAVGAQSDRMLNIPGEDLP
GSIAAVDFVGWYNAHPHFEQVSPDLSGARAVVIGNGNVALDVARILLTDPDVLARTDIAD
HALESLRPRGIQEVVIVGRRGPLQAAFTTLELRELADLDGVDVVIDPAELDGITDEDAAA
VGKVCKQNIKVLRGYADREPRPGHRRMVFRFLTSPIEIKGKRKVERIVLGRNELVSDGSG
RVAAKDTGEREELPAQLVVRSVGYRGVPTPGLPFDDQSGTIPNVGGRINGSPNEYVVGWI
KRGPTGVIGTNKKDAQDTVDTLIKNLGNAKEGAECKSFPEDHADQVADWLAARQPKLVTS
AHWQVIDAFERAAGEPHGRPRVKLASLAELLRIGLG*

MTSP32
VTNPPWTVDVVVVGAGFAGLAAARELTRQGHEVLVFEGRDRVGGRSLTGRVAGVPADMGG
SFIGPTQDAVLALATELGIPTTPTHRDGRNVIQWRGSARSYRGTIPKLSLTGLIDIGRLR
WQFERIARGVPVAAPWDARRARELDDVSLGEWLRLVRATSSSRNLMAIMTRVTWGCEPDD
VSMLHAARYVRAAGGLDRLLDVKNGAQQDRVPGGTQQIAQAAAAQLGARVLLNAAVRRID
RHGAGVTVTSDQGQAEAGFVIVAIPPAHRVAIEFDPPLPPEYQQLAHHWPQGRLSKAYAA
YSTPFWRASGYSGQALSDEAPVFITFDVSPHADGPGILMGFVDARGFDSLPIEERRRDAL
RCFASLFGDEALDPLDYVDYRWGTEEFAPGGPTAAVPPGSWTKYGHWLREPVGPIHWAST
ETADEWTGYFDGAVRSGQRAAAEVAALL*

MTSP33
MKGTKLAVVVGMTVAAVSLAAPAQADDYDAPFNNTIHRFGIYGPQDYNAWLAKISCERLS
RGVDGDAYKSATFLQRNLPRGTTQGQAFQFLGAAIDHYCPEHVGVLQRAGTR*

MTSP34
MKALVAVSAVAVVALLGVSSAQADPEADPGAGEANYGGPPSSPRLVDHTEWAQWGSLPSL
RVYPSQVGRTASRRLGMAAADAAWAEVLALSPEADTAGMRAQFICHWQYAEIRQPGKPSW
NLEPWRPVVDDSEMLASGCNPGSPEESF*

MTSP35
MSGRHRKPTTSNVSVAKIAFTGAVLGGGGIAMAAQATAATDGEWDQVARCESGGNWSINT
GNGYLGGLQFTQSTWAAHGGGEFAPSAQLASREQQIAVGERVLATQGRGAWPVCGRGLSN
ATPREVLPASAAMDAPLDAAAVNGEPAPLAPPPADPAPPVELAANDLPAPLGEPLPAAPA
DPAPPADLAPPAPADVAPPVELAVNDLPAPLGEPLPAAPADPAPPADLAPPAPADLAPPA
PADLAPPAPADLAPPVELAVNDLPAPLGEPLPAAPAELAPPADLAPASADLAPPAPADLA
PPAPAELAPPAPADLAPPAAVNEQTAPGDQPATAPGGPVGLATDLELPEPDPQPADAPPP
GDVTEAPAETPQVSNIAYTKKLWQAIRAQDVCGNDALDSLAQPYVIG*

MTSP36
MSGHRKKAMLALAAASLAATLAPNAVAAAEPSWNGQYLVTLSANAKTGTSMAANRPEYPH
KANYTFSSRCASDVCIATVVDAPPPKNEFIPRPIEYTWNGTQWVREISWQWDCLLPDGTI
EYAPAKSITAYTPGQYGILTGVFHTDIASGTCKGNVDMPVSAKPIVG*

FIG. 1 (continued)

MTSP37
MRYLIATAVLVAVVLVGWPAAGAPPSCAGLGGTVQAGQICHVHASGPKYMLDMTFPVDYP
DQQALTDYITQNRDGFVNVAQGSPLRDQPYQMDATSEQHSSGQPPQATRSVVLKFFQDLG
GAHPSTWYKAFNYNLATSQPITFDTLFVPGTTPLDSIYPIVQRELARQTGFGAAILPSTG
LDPAHYQNFAITDDSLIFYFAQGELLPSFVGACQAQVPRSAIPPLAI*

MTSP38
LKNARTTLIAAAIAGTLVTTSPAGIANADDAGLDPNAAAGPDAVGFDPNLPPAPDAAPVD
TPPAPEDAGFDPNLPPPLAPDFLSPPAEEAPPVPVAYSVNWDAIAQCESGGNWSINTGNG
YYGGLRFTAGTWRANGGSGSAANASREEQIRVAENVLRSQGIRAWPVCGRRG*

MTSP39
MSTIFDIRSLRLPKLSAKVVVVGGLVVVLAVVAAAAGARLYRKLTTTTVVAYFSEALALY
PGDKVQIMGVRVGSIDKIEPAGDKMRVTLHYSNKYQVPATATASILNPSLVASRTIQLSP
PYTGGPVLQDGAVIPIERTQVPVEWDQLRDSINGILRQLGPTERQPKGPFGDLIESAADN
LAGKGRQLNETLNSLSQALTALNEGRGDFVAITRSLALFVSALYQNDQQFVALNENLAEF
TDWFTKSDHDLADTVERIDDVLGTVRKFVSDNRSVLAADVNNLADATTTLVQPEPRDGLE
TALHVLPTYASNFNNLYYPLHSSLVGQFVFPNFANPIQLICSAIQAGSRLGYQESAELCA
QYLAPVLDALKFNYLPFGSNPFSSAATLPKEVAYSEERLRPPPGYKDTTVPGIFSRDTPF
SHGNHEPGWVVAPGMQGMQVQPFTANMLTPESLAELLGGPDIAPPPPGTNLPGPPNAYDE
SNPLPPPWYPQPASLPAAGATGQPGPGQ*

MTSP40
MKRSMKSGSFAIGLAMMLAPMVAAPGLAAADPATRPVDYQQITDVVIARGLSQRGVPFSW
AGGGISGPTRGTGTGINTVGFDASGLIQYAYAGAGLKLPRSSGQMYKVGQKVLPQQARKG
DLIFYGPEGTQSVALYLGKGQMLEVGDVVQVSPVRTNGMTPYLVRVLGTQPTPVQQAPVQ
PAPVQQAPVQQAPVQQAPVQQAPVQQAPVQQAPVQQAPVQQAPVQPPPFGTARSR*

MTSP41
MFTRRFAASMVGTTLTAATLGLAALGFAGTASASSTDEAFLAQLQADGITPPSAARAIKD
AHAVCDALDEGHSAKAVIKAVAKATGLSAKGAKTFAVDAASAYCPQYVTSS*

MTSP42
MAAMWRRRPLSSALLSFGLLLGGLPLAAPPLAGATEEPGAGQTPGAPVVAPQQSWNSCRE
FIADTSEIRTARCATVSVPVDYDQPGGTQAKLAVIRVPATGQRFGALLVNPGGPGASAVD
MVAAMAPAIADTDILRHFDLVGFDPRGVGHSTPALRCRTDAEFDAYRRDPMADYSPAGVT
HVEQVYRQLAQDCVDRMGFSFLANIGTASVARDMDMVRQALGDDQINYLGYSYGTELGTA
YLERFGTHVRAMVLDGAIDPAVSPIEESISQMAGFQTAFNDYAADCARSPACPLGTDSAQ
WVNRYHALVDPLVQKPGKTSDPRGLSYADATTGTINALYSPQRWKYLTSGLLGLQRGSDA
GDLLVLADDYDGRDADGHYSNDQDAFNAVRCVDAPTPADPAAWVAADQRIRQVAPFLSYG
QFTGSAPRDLCALWPVPATSTPHPAAPAGAGKVVVVSTTHDPATPYQSGVDLARQLGAPL
ITFDGTQHTAVFDGNQCVDSAVMHYFLDGTLPPTSLRCAP*

FIG. 1 (continued)

MTSP43
MKTGTATTRRRLLAVLIALALPGAAVALLAEPSATGASDPCAASEVARTVGSVAKSMGDY
LDSHPETNQVMTAVLQQQVGPGSVASLKAHFEANPKVASDLHALSQPLTDLSTRCSLPIS
GLQAIGLMQAVQGARR*

MTSP44
MSRLSSILRAGAAFLVLGIAAATFPQSAAADSTEDFPIPRRMIATTCDAEQYLAAVRDTS
PVYYQRYMIDFNNHANLQQATINKAHWFFSLSPAERRDYSEHFYNGDPLTFAWVNHMKIF
FNNKGVVAKGTEVCNGYPAGDMSVWNWA*

MTSP45
VTKRTITPMTSMGDLLGPEPILLPGDSDAEAELLANESPSIVAAAHPSASVAWAVLAEGA
LADDKTVTAYAYARTGYHRGLDQLRRHGWKGFGPVPYSHQPNRGFLRCVAALARAAAAIG
ETDEYGRCLDLLDDCDPAARPALGL*

MTSP46
VIIPDINLLLYAVITGFPQHRRAHAWWQDTVNGHTRIGLTYPALFGFLRIATSARVLAAP
LPTADAIAYVREWLSQPNVDLLTAGPRHLDIALGLLDKLGTASHLTTDVQLAAYGIEYDA
EIHSSDTDFARFADLKWTDPLRE*

MTSP47
LTDPRHTVRIAVGATALGVSALGATLPACSAHSGPGSPPSAPSAPAAATVMVEGHTHTIS
GVVECRTSPAVRTATPSESGTQTTRVNAHDDSASVTLSLSDSTPPDVNGFGISLKIGSVD
YQMPYQPVQSPTQVEATRQGKSYTLTGTGHAVIPGQTGMRELPFGVHVTCP*

FIG. 2

mtsp1
atgaatcgcatcgtgcagttcggagtttccgccgtggccgcggcggcgat
cggcatcggagccgggtcggggatcgcggcggcgttcgacggcgaggacg
aggtgaccggcccccgacgccgaccgcgcgcgcgccgccgcggtgcaggcg
gtcccgggcggcaccgccggagaagtcgagaccgagaccggcgaaggcgc
cgccgcctacggcgtgctggtcacccggcccgacggcacccgtgtcgagg
tccacctggaccgggatttccgggttctggacaccgaaccggccgacggg
gacggcggttag

mtsp2
atgaggctgtcgttgaccgcattgagcgccggtgtaggcgccgtggcaat
gtcgttgaccgtcggggccggggtcgcctccgcagatcccgtggacgcgg
tcattaacaccacctgcaattacgggcaggtagtagctgcgctcaacgcg
acggatccgggggctgccgcacagttcaacgcctcaccggtggcgcagtc
ctatttgcgcaatttcctcgccgcaccgccacctcagcgcgctgccatgg
ccgcgcaattgcaagctgtgccgggggcggcacagtacatcggccttgtc
gagtcggttgccggctcctgcaacaactattaa

mtsp3
atgttcaccggcatcgctagccatgccggcgccctgggtgccgccttagt
ggtgctgatcggcgccgcaattctgcacgacggcccagcagcggccgacc
caaaccaagacgatcggtttctggcgctgctcgagaaaaaggaaatcccc
gccgtcgcgaatgtgcctcgcgtcatcgacgcggcccacaaagtgtgtcg
caaactcgatggcggcatgccggtgaacgacattgtggacgggttacgca
acgatgcctacaacatagacccggtcatgcgcctctaccctgtccgcctc
acgacgaccatgacccgatttatcagtgcggcagtggagatctactgccc
gaaccatcacagcaagatggcgttcgccatggccaatttcgagccgggat
cgaatgaaccgacgcatcgcgttgcggcgtccacgcgcagcgcggtcaac
tcgggaagcgacctgcgggcgtcggtgtcggacatgaccatcatgtcgcc
gggatggcgggaaccgacgggtgcgatgcttgcctcggtgctcggagcgg
ttcgcgcgggggatcccctgataccgaatccgccgccgattccggtaccg
ccgccggcggcgcagaccctgattccaccccccgccgatcgtggcaccgcc
gccaccgcgaccagcgccgccgcaacagccgccgcccccgccgccagagg
ttgagccgcctgctggtgttccgcagtccgggggcgctgccggcagtggc
ggcgccggcagcggtggtggcggcggtggtgacggaccggtagagccgtc
gcctgcacgacccatgccgccgggctttatcaggctcgcgccgtga

mtsp4
atgacgcggctgataccgggttgcacgctcgtcgggctgatgctgacgtt
actgcccgcgcccacctcggcggccgggagcaacaccgccaccaccctgt
tcccggtcgacgaggtcacccagctggagacgcacaccttcctcgattgc
caccccaacggcagctgcgacttcgtcgctggagcaaatctgcgcacacc
cgacggcccgacgggctttccgcccgggctgtgggcgcgccaaaccaccg
agatccgttcgacgaaccggttggcctatctggacgcgcacgccaccagc
cagttcgaacgggtaatgaaggcgggcggatccgacgtgatcaccaccgt
ctacttcggcgagggtccgccggacaaataccagaccaccggggtcatcg
actcgaccaattggtcgaccggtcaaccgatgaccgacgtcaacgtcatc
gtgtgtacacacatgcaggtggtctacccgggggtcaacctcacctcgcc
cagcacctgcgcgcaagccaacttttcctag

FIG. 2 (continued)

mtsp5

atggttttaagaagtaggaaaagcacgctcggcgttgtcgtgtgcttagc
gctggtgctcggtgggccgctcaacggttgcagcagcagcgcgagccacc
gcggtccactgaacgcaatgggaagtccggccataccgtcgacggcgcag
gagatacccaacccgttgcgcggtcagtacgaagacctcatggaaccgct
gtttccgcaggggaaccccgcgcagcaacgctatccgccttggcccgcgt
cctacgacgcgagtttgcgagtctcctggcggcagctgcagcctacggat
ccgcgcactctgcccccggatgctccggacgaccgcaagtacgacttcag
cgtgatcgacaacgcgttgaccaggctcgccgaccgcggcatgcggctga
cgctgcgggtgtacgcctacagctcgtgctgcaaggcttcctatccggac
ggcactaacatcgcgattcccgactgggagcgcgctatcgccagcaccaa
caccagttatccagggccggcgaccgatccctcgaccggggtggtgcagg
tggtgccgaatttcaacgattcgacctatcttaacgattttgcgcagttg
ctcgccgcgcttggtcgccgctacgacggtgacgagcgcctcagcgtgtt
cgagttctccgggtacggggacttcagcgaaaatcacgtcgcatacctgc
gcgacacgctcggtgcgccgggtccgggcccggatgaaagcgtggcgacc
ctgggctattacagccagttccgtgatcagaacatcaccaccgcgtccat
caaacagctaatcgcggcgaacgtcagcgccttcccgcatacccaactgg
tgaccagtcccgctaatccggaaatcgtgcgagaactgttcgccgacgag
gtcaccaacaagcttgccgcgccggtgggtgtccgctcggattgcctggg
cgtcgacgcgccgttgccggcctgggccgagtccagcacttcgcactatg
tgcagaccaaagacccggtggtcgccgcgctgcggcagcggctggcaacg
gcgccggtgatcaccgagtggtgcgagttgccgaccggcagttcgccgcg
ggcttactacgagaagggcctgcgcgacgtcatcaggtatcacgtgtcga
tgacgtcgagcgttaacttccccgaccagacggcgacctcgccgatggac
cccgcgttgtacctggtgtgggcgcaagctaacgccgccgcaggctatcg
gtactcggtcgaagcgcagccggggtcgcaagcgctagcgggcaaggtcg
cgacgatctcggtcacctggaccaactacggcgctgctgccgccaccgaa
aagtgggtgcccggctaccggctggtggattccaccggacaggtggttcg
gacgctgccggcagcggtggacctgaagacgctggtctccgaccagcgcg
gcgatcgcagcagcgaccagccgacaccggcgtcggtcgccgagacggtt
cgcgttgatctgtccggcttgcccgcgggccactacacgctgcgggccgc
gatcgactggcaacagcacaaaccgaacggctcccatgtggtgaactatc
cgcccatgctgttgtcccgcgacggccgcgacgattccgggttttatccc
gtcgccacgctcgacatcccacgcgacgcgcagaccgcggtcaacgcttc
gtag

mtsp6

atgagccgactcctagctttgctgtgcgctgcggtatgcacgggctgcgt
tgctgtggttctcgcgccagtgagcctggccgtcgtcaacccgtggttcg
cgaactcggtcggcaatgccactcaggtggtttcggtggtgggaaccggc
ggttcgacggccaagatggatgtctaccaacgcaccgccgccggctggca
gccgctcaagaccggtatcaccacccatatcggttcggcgggcatggcgc
cggaagccaagagcggatatccggccactccgatgggggtttacagcctg
gactccgcttttggcaccgcgccgaatcccggtggcgggttgccgtatac
ccaagtcggacccaatcactggtggagtggcgacgacaatagccccacct
ttaactccatgcaggtctgtcagaagtcccagtgcccgttcagcacggcc
gacagcgagaacctgcaaatcccgcagtacaagcattcggtcgtgatggg
cgtcaacaaggccaaggtcccaggcaaaggctccgcgttcttctttcaca
ccaccgacggcgggcccaccgcgggttgtgtggcgatcgacgatgccacg

FIG. 2 (continued)

ctggtgcagatcatccgttggctgcggcctggtgcggtgatcgcgatcgc
caagtaa

mtsp7
atgattcgcgaactggtcaccaccgctgcgatcacgggtgccgcgatcgg
tggggcgccagtcgcgggcgcagacccgcagcgttatgacggcgatgtgc
cggggatgaactatgacgcttcgctgggcgccccatgctccagctgggag
cgcttcatttttggacgaggcccctccggtcaggccgaagcctgtcattt
tccgcctcctaaccagttcccgccggccgaaaccggctactgggtgatct
cctacccgctatacggcgtccagcaggtcggtgcgccgtgtccgaagccg
caggcggccgcgcagtctccggatgggttgccgatgctgtgtctgggagc
ccgtggatggcagccgggatggtttaccggggccgggttcttccctccgg
agccataa

mtsp8
atgggtgaattacggttggtgggcggtgtgctccgggtccttgtcgtggt
cggtgcggtgttcgatgtggcggtgctaaacgccggtgcggctagtgccg
acggcccggtccagctgaagagccgattgggcgatgtttgcctggacgcc
ccgagtgggagctggttcagcccgctggtgatcaacccctgcaatgggac
cgactttcagcgctggaatctcaccgatgaccggcaggtcgagagcgtgg
ccttccccggggaatgcgtgaatatcggaaatgctttgtgggcgcgcctg
cagccctgtgtgaactggatcagccagcactggactgtccagcccgacgg
cctggtcaagagtgatcttgatgcctgcctcacggttctcggcggtccgg
atcctgggacctgggtgtccacccgctggtgcgaccccaatgcacccgac
caacagtgggatagcgtgccgtaa

mtsp9
atgccggccatgaccgcccgttcggtggtactcagcgtgctgctcggtgc
tcatcccgcgtgggccaccgcaagcgaattgatccagctgacagcggatt
tcggtatcaaggagacgacgttgcgggtcgcgctgacccgcatggtcggt
gccggggatctggtccggtccgcggacggctaccggctctcggatcggtt
gctggcccgccagcgccgacaagatgaggccatgcgcccacggacccgcg
cttggcacggaaactggcacatgctgattgtcaccagcatcggcaccgat
gctcgtacccgggccgcactgcgaacctgcatgcaccacaagcgtttcgg
tgaattgcgggaaggggtgtggatgcggccggacaatctcgacctcgact
tggagtccgacgttgcggcccgggttaggatgctgacggcccgcgacgag
gcccccgccgacttggccgggcagctgtgggatctgtcggggtggaccga
ggccggccaccggttgctcggcgacatggcagcggccaccgacatgcccg
ggcgatttgtggtggctgcggcgatggtgcgccacctgctcaccgatccg
atgttgcccgctgaactgttgcccgccgactggccgggcgccgggttacg
ggcggcgtaccacgacttcgccactgcaatggcgaaacgacgcgatgcaa
ctcaactcctggaggtgacatga

mtsp10
gtgccggccggcgtcggtaacgcatccggtagcgttttagatatgacgtc
cgtgcgcacagtgccaagcgccgtcgcgctggtgacgtttgccggagccg

FIG. 2 (continued)

cgctcagcggggtcatcccggcgattgcccgcgcggatccggtcgggcat
caggtgacctacaccgtcacgaccaccagcgacctgatggccaacattcg
gtacatgagcgccgatccgcccagcatggcggctttcaatgccgattcat
cgaagtacatgattaccttgcacactccgatcgctggcggtcagccgctg
gtctataccgccacgctggcaaacccgagccagtgggcgatcgtcaccgc
cagcggcggcctgcgggtcaatccggagttccactgcgagattgttgtag
acggccaggtggtggtgtcgcaggacggcggcagcggcgtgcagtgctcg
actcgtccctggtaa

mtsp11
atgacgaccagcaaaatcgccaccgccttcaagaccgccaccttcgcgct
ggccgccggtgccgttgcactgggattggccagccccgccgacgcagcgg
cgggcaccatgtatggcgacccggcagccgccgccaagtactggcgccag
cagacatacgacgactgcgtcctgatgtcggccgcggacgtgatcggtca
agtgaccggcagggagccttccgagcgcgccatcatcaaagtggcccagt
cgacacccagcgtcgtgcaccccgggtccatctacacaaagccggccgac
gccgagcacccgaactcgggaatgggtaccagcgtggccgacataccgac
gctgctggcgcattacggcgtcgacgccgttatcaccgacgaggaccacg
ccacagccaccggagtcgccaccggcatggccgccctcgagcagtatctg
ggcagcgggcacgccgtgatcgtcagcatcaacgccgagatgatctgggg
ccagcccgtcgaggaaaccgacagtgccggcaacccgcggtctgaccacg
ccgtggtggtgaccggtgtcgataccgaaaacggcattgttcacctcaac
gacagcggtacccccacgggccgcgacgagcagatcccgatggaaacctt
cgtcgaggcgtgggccaccagccacgacttcatggccgtcaccacctga

mtsp12
atgggagtcattgcccgcgttgtcggtgtcgccgcgtgcggtttgtccct
ggccgtgctggccgccgcgcccaccgcgggcgcggaacccaccggcgcgc
tgcccccgatgacatccagcggcagcggaccggtcatcggcgacggtgac
gccgcgctgcgacagcggatctcacagcagctgtttagcttcggagatcc
caccgtccaggaggttgacggctcggacgcggctcaattcatcacggccg
cagccgctgtcgcggaccgcgatgtggcgtcggtgttcttgccgctgcag
cgggtgttgggctgccaacagaacacagccggctcgggggccggcttcgg
ggcgcgcgcctaccggcgaaccgacgggcaatggggaggcgcgatgctgg
tcgtcgccaagagcaccgtttccgacgtcgacgccctcaaggcctgcgtc
aagtccggttggcgcaaggccacggcgggcacgccgacttcgatgtgcaa
caacggttggacctacccgccgttcgccgacacccgccgcggcgaagagg
gctatttcgtcttgctggccggcacggcctcggacttctgcagtgcgccc
aacgcgaactaccgaaccaccgcgagctcatggccgggctag

mtsp13
atgcgcttgaagccagccccatctcctgctgcagcctttgccgtcgccgg
cctgatcctcgcaggctgggccggatccgtgggcctcgccggcgccgatc
cggagccggcaccgacaccgaagacggcaattgatagcgacggcacctat
gcggtggggattgacatcgctcccggcacgtacagctccgcgggacccgt
cggcgacggcacctgctattggaagcggatgggtaaccccgatggcgcgc
tcatcgataacgcactcagcaagaaaccacaggtagtgacgattgagccg
accgacaaggcgttcaagacgcacggctgccaaccctggcagaacacggg
cagcgaaggcgctgcccctgccggagttcctggacctgaagcgggggccc

FIG. 2 (continued)

aactacaaaatcagctcggcatcctcaacggcttactcggaccgactgga
gggcgagtgcctcagccctaa

mtsp14
atgatcacaaacctccgacgccgaaccgcgatggcagccgccggcctagg
ggctgctctcgggctgggcatcctgctggttccgacggtggacgcccatc
tcgccaacggttcgatgtcggaagtcatgatgtcggaaattgccgggttg
cctatccctccgattatccattacggggcgattgcctatgcccccagcgg
cgcgtcgggcaaagcgtggcaccagcgcacaccggcgcgagcagagcaag
tcgcactagaaaagtgcggtgacaagacttgcaaagtggttagtcgcttc
accaggtgcggcgcggtcgcctacaacggctcgaaataccaaggcggaac
cggactcacgcgccgcgcggcagaagacgacgccgtgaaccgactcgaag
gcgggcggatcgtcaactgggcgtgcaactaa

mtsp15
gtgacggtgctgctcgacgccaacgtgctgatcgcattggtggtcgccga
gcatgtgcatcatgatgccgcagcggactggctcatggcgtccgacaccg
gatttgcgacctgcccgatgacacaaggaagcctggttcgattcctggtg
cgctcgggacagtccgcggcggcggctcgggatgtcgtcagtgcggtcca
gtgcacgagccgccacgaattctggcccgatgcactctctttcgccggtg
tcgaggtcgctggtgtggttgggcaccggcaggtgaccgatgcctacctt
gcccagctcgcgcgaagccacgacgggcagttggcgacgctcgacagcgg
cttagcacacctgcacggcgacgtcgcggtactcattccaacgaccacct
ga

mtsp16
gtgcagcgccaatcattgatgccccagcagacccttgccgccggcgtttt
cgtgggtgcgctgctatgcggtgtcgtgacggcggcggtgccaccacacg
cacgcgccgacgtggtcgcctatctggtcaacgtgacggtacgcccgggc
tacaacttcgccaacgccgacgccgcgttgagttacggacatggcctctg
cgagaaggtgtctcggggccgcccttacgcacagatcatcgccgacgtca
aggctgatttcgacacccgcgaccaataccaggcctcgtatctgctcagc
caggctgtcaacgaactctgccccgcgctgatctggcagttgcgaaactc
cgcagtcgacaatcggcgctcgggctga

mtsp17
gtgcgttcgtatctattgcgtatcgagctggccgaccggccgggcagcct
tgggtcgctggcggtcgcgctcggctcggtgggcgccgacatcctctcgc
tcgacgtggtcgagcgcggcaacggctatgcgatcgacgacctggtggtc
gaactgcccccgggagcgatgcccgacacgctgatcactgctgccgaggc
gctgaacggcgtccgggtagacagcgtccgcccgcacaccggcctgttgg
aagcccaccgcgagctggaactgctcgatcatgtggccgcggctgagggc
gcgaccgcacggctccaggttctggtcaacgaggcccccgggtgctccg
ggtgagctggtgcacggtgttgcgcagttccggcggggagctgcaccgtc
tggccggcagcccaggtgcgccggagacccgggccaattcggcgccctgg
ctgccgatcgagcgggccgcggcgctggacggcggcgccgactgggtgcc
gcaagcctggcgcgacatggataccaccatggtcgcggctccattgggtg
acacgcacaccgcggtggtgctgggcaggccaggcccggaatttcgcccg

FIG. 2 (continued)

tcggaggtggcgcggttgggttatctagccggcatcgtggcgacgatgct
gcgctga

mtsp18
atgcctgacggggagcagagccagccaccggcccaagaagatgcggaaga
cgactcgcggcccgacgccgcggaggccgccgcggccgaacccaaatcat
cagccggtccgatgttctcgacctacggtatcgcctcgacactactcggc
gtgctatcggtcgccgcggtcgtgctgggtgcgatgatctggtccgcaca
ccgcgatgactccggcgagcgtacctacctgacccgggtcatgctgaccg
ccgctgaatggacggccgtgctgatcaacatgaacgccgacaacatcgat
gccagcctgcagcgactgcacgacggaacggtcggtcaactcaacaccga
cttcgacgctgtcgtgcagccctaccggcaggtggtggagaagttgcgga
cgcacagcagcggcaggatcgaggcggtagcgatcgatacggtgcaccgc
gagctggatacccagtccggtgccgcccgaccggtagtaaccacgaaatt
gccaccgtttgccactcgcaccgactcggtgctgctggtcgcgacgtcgg
tcagtgagaacgccggcgccaaaccccagaccgtgcactggaacttgcgg
ctcgatgtctccgatgtggacggcaagctgatgatctcccggttggagtc
gattcgatga

mtsp19
atgaagatggtgaaatcgatcgccgcaggtctgaccgccgcggctgcaat
cggcgccgctgcggccggtgtgacttcgatcatggctggcggcccggtcg
tataccagatgcagccggtcgtcttcggcgcgccactgccgttggacccg
gcatccgcccctgacgtcccgaccgccgcccagttgaccagcctgctcaa
cagcctcgccgatcccaacgtgtcgtttgcgaacaagggcagtctggtcg
agggcggcatcgggggcaccgaggcgcgcatcgccgaccacaagctgaag
aaggccgccgagcacggggatctgccgctgtcgttcagcgtgacgaacat
ccagccggcggccgccggttcggccaccgccgacgtttccgtctcgggtc
cgaagctctcgtcgccggtcacgcagaacgtcacgttcgtgaatcaaggc
ggctggatgctgtcacgcgcatcggcgatggagttgctgcaggccgcagg
gaactga

mtsp20
atgaacctacggcgccatcagaccctgacgctgcgactgctggcggcatc
cgcgggcattctcagcgccgcggccttcgccgcgccagcacaggcaaacc
ccgtcgacgacgcgttcatcgccgcgctgaacaatgccggcgtcaactac
ggcgatccggtcgacgccaaagcgctgggtcagtccgtctgcccgatcct
ggccgagcccggcgggtcgtttaacaccgcggtagccagcgttgtggcgc
gcgcccaaggcatgtcccaggacatggcgcaaaccttcaccagtatcgcg
atttcgatgtactgcccctcggtgatggcagacgtcgccagcggcaacct
gccggccctgccagacatgccggggctgcccgggtcctag

mtsp21
atgagagttgtgtcaacgctactcagcattccgttgatgatcggcttggc
ggttccggcccacgcggggcccagcggtgacgacgcggtctttcttgcct
cgctagagcgggcaggcattacctacagccacccggatcaagccatagca
tcgggcaaggccgtatgcgcgttagtcgaaagcggcgaatcgggtcttca
ggtcgtcaacgagctgcggacccgcaatcccgggttttcgatggacggtt
gttgcaagttcgctgcgatctccgcgcatgtctattgcccccaccagatc
actaaaaccagcgtcagcgcgaaatag

FIG. 2 (continued)

mtsp22
atggcccgcacgcttgcgttgcgcgcatcggcgggactcgtcgcgggtat
ggcaatggccgcgatcacgctcgcacctggggcccgcgccgaaaccggtg
agcaattccccggggatggggtgtttctcgtgggaactgacattgcgcca
ggcacctaccgcacggaggggccgtcgaatccccttattttggtgttcgg
cagggtgtccgagctctcaacctgctcatggtcgacacacagcgcacccg
aggtgagcaatgagaacattgtcgacaccaacacctctatgggcccgatg
tcagtggtgatcccgccgaccgtggcagccttccagacgcataactgcaa
gctttggatgcggatctcatag

mtsp23
atgttatcgccgttatcgcctcgcattatcgcagcgttcaccactgcagt
cggcgccgccgccatcggacttgccgtcgccaccgccggcaccgccggcg
ccaacaccaaagacgaagccttcattgctcagatggagtccattggcgtc
accttctcctcaccgcaggtggccacccagcaagcccagctggtctgcaa
gaagctggccagcggcgaaaccggcaccgagatcgccgaggaggtcctca
gccaaaccaacctgaccactaagcaggcagcctacttcgtcgtcgacgca
accaaggcctactgcccgcaatacgccagccagctcacctag

mtsp24
atgacgacgatgattactcttcggcgacggttcgcggtggccgtcgccgg
cgtcgccactgccgccgcgacgaccgtcaccctggctcccgcaccagcaa
atgccgccgatgtctatggcgcaattgcctactccggcaacggctcgtgg
ggccgatcgtgggactacccaacccgggcggctgccgaagccaccgccgt
caagtcgtgtggctactccgactgcaaggtgctcaccagtttcaccgcct
gcggcgccgtcgccgccaacgatagggcataccagggaggagttggaccc
accttggccgccgccatgaaggacgccctgaccaagctcggcggcggcta
catcgacacctgggcctgcaactaa

mtsp25
atgacaccgggtttgcttactactgcgggtgctggccgaccacgtgacag
gtgcgccaggatcgtatgcacggtgttcatcgaaaccgccgttgtcgcga
ccatgtttgtcgcgttgttgggtctgtccaccatcagctcgaaagccgac
gacatcgattgggacgccatcgcgcaatgcgaatccggcggcaattgggc
ggccaacaccggtaacgggttatacggtggtctgcagatcagccaggcga
cgtgggattccaacggtggtgtcgggtcgccggcggccgcgagtccccag
caacagatcgaggtcgcagacaacattatgaaaacccaaggcccgggtgc
gtggccgaaatgtagttcttgtagtcagggagacgcaccgctgggctcgc
tcacccacatcctgacgttcctcgcggccgagactggaggttgttcgggg
agcagggacgattga

mtsp26
gtgcaaggagccgttgctggtctggtgtttctggccgtcctggtgatttt
cgccatcatcgtggtggccaagtcggtggcgctgatcccgcaggcggagg
ccgcggtgatcgagcggctgggtcgctatagtcgtacggtcagtgggcag
ttgacgctgttggtgccgttcatcgaccgcgtccgggctcgggtggacct
gcgcgagcgggtggtgtcgtttccgccgcaaccggtgatcaccgaggaca
acttgacgctgaacatcgacaccgtcgtctacttccaggtgaccgttccg

FIG. 2 (continued)

caggcggcggtgtacgagatcagcaattacatcgtcggggtcgaacagct
caccaccaccaccctgcgcaacgttgtcggcgggatgacgctggagcaga
cgttgacctcgcgtgaccagatcaacgcccagctgcgcggcgttctcgat
gaggcgaccggccgctggggtctgcgggtggcgcgggtggagctgcgcag
catcgatccgccgccgtcgattcaggcgtcgatggaaaagcagatgaagg
ccgaccgggagaagcgagcgatgattctgaccgccgaaggtacccgggag
gcggcgataaaacaggccgaggggcaaaagcaggcgcagatcctggccgc
cgagggcgccaagcaggccgcgatcttggctgctgaggccgatcggcagt
ctcggatgctgcgcgctcagggtgagcgcgccgcggcctacctgcaggcg
caagggcaggccaaggccatcgagaagacgttcgccgcgatcaaggctgg
ccggcccaccccggagatgctggcctaccaatacctgcagacgctgccgg
agatggcgcgtggggacgccaacaaggtatgggtggtgcccagcgacttc
aacgccgcactgcaggggttcaccaggctgctgggcaagccgggtgagga
cggggtgttccggttcgagccgtccccggtcgaagaccagcccaagcacg
cggccgacggtgacgacgccgaggtcgccggctggttctccaccgatacc
gacccgtcgatcgctcgggcggtggctacagccgaggcgatagcccgcaa
gccggtcgagggttcgctggggacgccccccaggttgactcaatag

mtsp27
ttgcagacggcgcacaggcgctttgccgcggcattcgcggccgtgctttt
ggccgttgtgtgcctacctgcgaacaccgcggcagccgacgacaagctac
cgctgggcggtggtgcgggcatcgtcgtcaacggggacaccatgtgcacc
ctaaccaccatcggccatgacaagaacggtgacctcatcggcttcacttc
cgcccactgtgggggcccgggcgcgcagatcgccgctgagggtgccgaga
acgcgggcccggtaggcatcatggtcgccggcaacgacggcctggactac
gcggtgatcaagttcgacccggccaaggtgaccccggtggccgtcttcaa
cgggtttgcgatcaacggcattggcccggacccgtcgttcggccagatcg
cctgcaagcagggccgcaccaccggtaactcgtgcggggttacctggggg
ccaggggagagtccgggcacccttgtgatgcaggtctgcggcggaccggg
cgactccggtgcgccggtgaccgtcgacaatctgctggtcgggatgatcc
acggcgcattcagcgacaatctgccgagttgcatcaccaaatacatcccg
ctgcacaccccggcggtggtgatgtcgatcaacgccgacctggccgacat
caacgccaagaaccggccgggcgcgggattcgtcccggtaccggcctga

mtsp28
atgcttatgcctgagatggatcgtcgccgaatgatgatgatggcggggtt
cggcgccctggctgccgcgcttcccgccccgacagcctgggccgacccgt
cccggccggccgcgccggctggtccgacaccggcgcccgccgcgccggct
gcggcaaccggtgggcttttgttccacgacgagttcgacgggccggccgg
ttcggtcccggacccgtccaagtggcaggtgtcgaaccaccggacgccca
tcaagaacccggtgggctttgaccggccccagtttttgggcagtaccgc
gacagtcgacagaacgtgttcctcgacggcaactccaatctcgtgctgcg
cgctacccgagagggcaacaggtatttcggtggcctggtccacggcctgt
ggcggggtggcatcgggaccacctgggaggcccggatcaagttcaactgc
ctggctccgggcatgtggcccgcctggtggttgtccaatgacgatcctgg
tcgcagcggcgaaatcgacctgatcgagtggtatggcaacgggacttggc
cgtcgggaaccaccgtgcacgccaacccggacggcaccgcattcgagacc
tgcccgatcggtgtggacggtggttggcacaactggcgcgtcacgtggaa
tccgagcggcatgtacttctggctggattacgccgacggcattgagccct
acttctcggttccggcgaccggaatcgaagacctcaacgagcccatccgc

FIG. 2 (continued)

gagtggccgttcaacgaccccggctacaaggtgtttccggtgttgaacct
tgcggttggcggttctggtggcggcgatcccgcgacgggttcctatccac
aggagatgctcgtcgactgggtgcgcgtcttttaa

mtsp29
gtgcaccgtcgaacggccctgaagctcccgctgctgctggcggcaggcac
ggtgctgggccaagcgccgcgggccgccgccgaagaaccaggccggtggt
cggccgaccgcgcacatcgctggtatcaagcgcacggctggctcgtcggt
gcaaactacatcacctcgaacgccatcaaccagctcgagatgttccagcc
aggcacatacgatccccggcgcatcgacaacgagctgggccttgcgcggt
ttcacgggttcaacaccgtgcgagtcttcctccacgacctgctgtgggcc
caagacgcgcccggtttccaaacccggctcgcgcagttcgtcgccatcgc
ggcgcgataccacatcaaaccgctctttgtcctgttcgactcctgctggg
acccgctccccagaccgggtcggcagcgggcgccaagggctggggtgcac
aactccgggtgggtgcaaagtccgggtgctgaacgcctcgatgaccgccg
ctatgccagcacgctgtacaactacgtcacgggtgtgttgggccaattcc
gcaacgacgatcgcgtgttgggttgggacctgtggaatgaacccgacaat
cccgcgcgcgtgtatcgcaaggtggaaaggaaagacaagctcgagcgcgt
cgcggagctcctcccccaagtgttccgatgggcccgcacggtcgatccgg
ttcaaccgctgaccagtggtgtctggcaagggaattggggagatcccgga
cgccgcagcaccatcagcgccattcaactcgacaacgccgacgtgatcac
cttccacagttacgccgcgccggccgaattcgagggccgcatcgctgagc
tcgctccgttgcagcggccaatcctgtgcaccgagtacctggcgcggtcc
caaggcagcactgtcgagggaatcctgccgattgctaagcggcacaacgt
tggtgcgttcaattggggtttggtggcgggaaagactcagacctatttgc
cgtgggattcgtgggatcacccctaccgcgcgccccgaaggtgtggttt
cacgacctgctacaccccaacggccggccgtatcgggacggcgaagttca
aacgattcggaagctgaacgggatgccgagccaggactag

mtsp30
gtgtccacgtacggctggcgcgcctacgccctgccggttctgatggtgct
gaccacggtggtggtgtaccagacggtgaccgggacgagcacgccaaggc
ccgcggcggcccagaccgtccgggactcgccggccattggtgtggtgggg
accgcgatcctcgacgcaccgcctcgcggtcttgcagtgttcgatgccaa
tctgccggccgggacgctgccggatggcggcccgttcaccgaggctggtg
acaagacctggcgtgtcgttccgggcactactccccaggtcggtcaaggc
accgtcaaagtgttcaggtataccgtcgagatcgagaacggtcttgatcc
cacaatgtacggcggtgacaacgcattcgcccagatggtcgaccagacgt
tgaccaatcccaagggctggacccacaatccgcaattcgcgttcgtgcgg
atcgacagcggaaaacccgacttccggatttcgctggtgtcgccgacgac
agtgcgcgggggtgtggctacgaattccggctcgagacgtcctgctaca
acccgtcgttcggcggcatggatcgccaatcgcgggtgttcatcaacgag
gcgcgctgggtacgcggagccgttccattcgaaggtgacgtaggttccta
tcggcaatatgtgatcaaccacgaggtcggtcatgccatcggttacctgc
gccacgagccgtgcgaccaacaaggcggtctggctccggtaatgatgcag
cagacgttttccacctccaatgacgacgcggccaagtttgaccccgactt
cgttaaggcggatggaaagacctgccgattcaatccctggccctacccga
ttccctaa

FIG. 2 (continued)

mtsp31

atgcgtccctattacatcgccatcgtgggctccgggccgtcggcgttctt
cgccgcggcatccttgctgaaggccgccgacacgaccgaggacctcgaca
tggccgtcgacatgctggagatgttgccgactccctgggggctggtgcgc
tccggggtcgcgccggatcaccccaagatcaagtcgatcagcaagcaatt
cgaaaagacggccgaggacccccgcttccgcttcttcggcaatgtggtcg
tcggcgaacacgtccagcccggcgagctctccgagcgctacgacgccgtg
atctacgccgtcggcgcgcagtccgatcgcatgttgaacatccccggtga
ggacctgccgggcagtatcgccgccgtcgatttcgtcggctggtacaacg
cacatccacacttcgagcaggtatcacccgatctgtcgggcgcccgggcc
gtagttatcggcaatggaaacgtcgcgctagacgtggcacggattctgct
caccgatcccgacgtgttggcacgcaccgatatcgccgatcacgctttgg
aatcgctacgcccacgcggtatccaggaggtggtgatcgtcgggcgccga
ggtccgctgcaggccgcgttcaccacgttggagttgcgcgagctggccga
cctcgacggggttgacgtggtgatcgatccggcggagctggacggcatta
ccgacgaggacgcggccgcggtgggcaaggtctgcaagcagaacatcaag
gtgctgcgtggctatgcggaccgcgaaccccgcccgggacaccgccgcat
ggtgttccggttcttgacctctccgatcgagatcaagggcaagcgcaaag
tggagcggatcgtgctgggccgcaacgagctggtctccgacggcagcggg
cgagtggcggccaaggacaccggcgagcgcgaggagctgccagctcagct
ggtcgtgcggtcggtcggctaccgcggggtgcccacgcccgggctgccgt
tcgacgaccagagcgggaccatccccaacgtcggcggccgaatcaacggc
agccccaacgaatacgtcgtcgggtggatcaagcgcgggccgaccggggt
gatcgggaccaacaagaaggacgcccaagacaccgtcgacaccttgatca
agaatcttggcaacgccaaggagggcgccgagtgcaagagctttccggaa
gatcatgccgaccaggtggccgactggctagcagcacgccagccgaagct
ggtcacgtcggcccactggcaggtgatcgacgctttcgagcgggccgccg
gcgagccgcacgggcgtccccgggtcaagttggccagcctggccgagctg
ttgcggattgggctcggctga

mtsp32

gtgacaaacccaccgtggactgtcgatgttgtcgtggtgggcgcgggctt
cgccgggctggccgcggcccgcgagctgacgcgacagggtcacgaggtgc
tggtgttcgaaggccgcgatcgggtgggcggccgctcgttaaccggtcgc
gtggcaggggtgcccgcggatatgggcggctcgttcatcggcccgaccca
agacgccgtgctggcgttggccaccgagctggggatcccgacaaccccga
cccaccgcgacggccgaaacgtcatccagtggcggggatcggcacgcagc
tatcgtggcaccatccccaagctgtcgctgaccgggctcatcgacatcgg
ccggttgcgttggcaattcgagcgaattgcccgcggcgttccggtggccg
ccccctgggatgcgcggcgcgcgcgtgaactcgacgacgtgtcgctcggg
gagtggttgcgcttggtgcgcgccacatcgtcctcgcggaacctgatggc
catcatgacccgggtgacctggggttgtgagcccgacgatgtctcgatgc
tgcacgccgcccgctacgtacgcgcggccggcggcctggaccggctgctc
gacgtcaaaaatggtgcccagcaggaccgtgtgccgggggggacacagca
gatcgcccaggcggccgccgcccaactcggcgcacgcgtcctgctcaacg
ccgcggtgcgtcgcatcgaccggcacggagcgggtgtgacggtcacgtcc
gatcagggtcaggccgaggccgggttcgtcatcgtcgccattccaccggc
ccatcgcgtggccatcgagttcgatcccccgctgccgccggaatatcagc
agctcgcccaccattggccgcagggccggctgagcaaggcctacgcggcc
tattcgacgccgttctggcgggccagcgggtattccggccaggcgctgtc
cgatgaggcgccggtgttcatcaccttcgacgtcagtccgcacgccgacg

FIG. 2 (continued)

ggccaggcattctgatggggttcgtcgatgctcgcgggttcgactcgcta
cccatcgaagagcgccgccgcgatgcattgcgctgctttgcgtcgctgtt
cggcgacgaagcgctcgaccccccttgattatgttgactatcgttggggta
cagaggaattcgcgccgggtggtccgaccgcggcggtaccgccggggtcg
tggacgaaatacggtcactggttacgtgagccggtcggtccgattcactg
ggcgagcactgagaccgcggacgaatggaccgggtatttcgacggcgccg
tcagatccggtcagcgtgccgccgccgaggtcgccgccctgctatga

mtsp33
atgaagggaacaaagctggctgttgtcgtcggcatgacggtggctgccgt
tagtttggcagcgccggcgcaggccgacgactacgacgccccttcaaca
acacgatccatcgcttcgggatctacggcccgcaggactacaacgcttgg
cttgccaagatcagctgcgaacggctgagcagaggcgttgacggcgatgc
gtacaagtcggccactttcctgcaacgcaacctgccgcgcggaaccaccc
agggccaagcgtttcagttcctgggcgccgcgatcgatcactactgccct
gagcatgtgggcgtcctgcaacgggctggcacccgctaa

mtsp34
atgaaagccctggtggccgtgtcggcggtggccgtcgtcgcactgctcgg
tgtatcttccgcccaagctgatcccgaggcggatcccggcgcaggtgagg
ccaactatggtggcccccaagttccccacgtcttgtcgatcacaccgaa
tgggcgcagtggggaagtctgcccagcctccgggtctacccgtcccaagt
tgggcgtacagcctcccgccgcctcgggatggccgctgccgacgcggcct
gggccgaggttctcgcgctgtcaccggaggccgacactgccggcatgcgc
gcgcagttcatctgccactggcagtacgccgaaatcagacaacccggcaa
acccagctggaacctcgagccgtggcggccggtcgtcgacgactcggaga
tgttggcttccggctgcaatccgggcagccctgaagagtcgttttag

mtsp35
atgagtggacgccaccgtaagcccaccacatccaacgtcagcgtcgccaa
gatcgcctttaccggcgcagtactcggtggcggcggcatcgccatggccg
ctcaggcgaccgcggccaccgacggggaatgggatcaggtggcccgctgc
gagtcgggcggcaactggtcgatcaacaccggcaacggttacctcggtgg
cttgcagttcactcaaagcacctgggccgcacatggtggcggcgagttcg
ccccgtcggctcagctggccagccgggagcagcagattgccgtcggtgag
cgggtgctggccacccagggtcgcggcgcctggccggtgtgcggccgcgg
gttatcgaacgcaacaccccgcgaagtgcttcccgcttcggcagcgatgg
acgctccgttggacgcggccgcggtcaacggcgaaccagcaccgctggcc
ccgccgcccgccgacccggcgccacccgtggaacttgccgctaacgacct
gcccgcaccgctgggtgaacccctcccggcagctcccgccgacccggcac
cacccgccgacctggcaccacccgcgcccgccgacgtcgcgccacccgtg
gaacttgccgtaaacgacctgcccgcaccgctgggtgaacccctcccggc
agctcccgccgacccggcaccacccgccgacctggcaccacccgcgcccg
ccgacctggcgccacccgcgcccgccgacctggcgccacccgcgcccgcc
gacctggcaccacccgtggaacttgccgtaaacgacctgcccgcgccgct
gggtgaacccctcccggcagctcccgccgaactggcgccacccgccgatc
tggcacccgcgtccgccgacctggcgccacccgcgcccgccgacctggcg
ccacccgcgcccgccgaactggcgccacccgcgcccgccgacctggcacc
acccgctgcggtgaacgagcaaaccgcgccgggcgatcagcccgccacag

FIG. 2 (continued)

ctccaggcggcccggttggccttgccaccgatttggaactccccgagccc
gacccccaaccagctgacgcaccgccgcccggcgacgtcaccgaggcgcc
cgccgaaacgccccaagtctcgaacatcgcctatacgaagaagctgtggc
aggcgattcgggcccaggacgtctgcggcaacgatgcgctggactcgctc
gcacagccgtacgtcatcggctga

mtsp36
atgtccggacaccgcaagaaggcaatgctcgccttggcggctgcgtcgct
ggcagcgacgctggccccgaacgcagtcgcggccgcagaaccgtcgtgga
acgggcagtacctcgtgacgttgtctgccaacgcgaaaaccggcaccagc
atggcggccaaccggccagagtatccacacaaagcgaactacacgttcag
ctcgcgctgcgcgtccgatgtctgcattgccaccgtggtcgacgctccgc
caccaaaaaacgagttcatcccgcggccaatcgaatacacctggaatggg
actcaatgggtacgggagatcagctggcaatgggactgcctgctacccga
cggcacaatcgaatatgccccagccaaatcgatcacggcctacacgcccg
gtcagtacggaatcctcaccggcgtctttcataccgatatcgccagcggc
acgtgtaaaggcaatgtcgacatgccagtgtcggccaaaccgatcgttgg
ctga

mtsp37
atgcgttatctgatagcgaccgcagtgctcgttgctgtggtcctggtggg
ctggccggcggctggtgcgccgccgtcatgcgccggcctgggcggcactg
tgcaggccggccagatctgccatgtgcacgcctcgggccctaagtacatg
ctggatatgacatttcctgtcgactatcccgaccagcaggcgctgaccga
ctacatcacgcaaaaccgcgacgggttcgtcaacgtcgcgcaggggtccc
cgctgcgagaccagccctaccaaatggacgccaccagcgaacagcacagc
tccggccagccgccgcaggccacccgcagcgtagtgctcaaattcttcca
ggacctcggtggggcacatccgtccacctggtacaaggccttcaactaca
acctcgcgacctcgcagcccatcaccttcgacacgttgttcgtgcccggc
accacgccactggacagcatctaccccatcgttcagcgcgagctggcacg
tcagaccggtttcggtgccgcgatattgccttcgaccggcctcgacccgg
ctcactaccagaactttgctatcaccgacgacagtctgattttctacttc
gcccagggtgagctgctgccgtcgtttgtcggcgcttgccaagcccaggt
gccgcgcagcgccattccgccgctggcaatctaa

mtsp38
ttgaagaacgcccgtacgacgctcatcgccgccgcgattgccgggacgtt
ggtgaccacgtcaccagccggtatcgccaatgccgacgacgcgggcttgg
acccaaacgccgcagccggcccggatgccgtgggctttgacccgaacctg
ccgccggccccggacgctgcacccgtcgatactccgccggctccggagga
cgcgggctttgatcccaacctccccccgccgctggccccggacttcctgt
ccccgcctgcggaggaagcgcctcccgtgcccgtggcctacagcgtgaac
tgggacgcgatcgcgcagtgcgagtccggtggaaactggtcgatcaacac
cggtaacggttactacggcggcctgcggttcaccgccggcacctggcgtg
ccaacggtggctcggggtccgcggccaacgcgagccgggaggagcagatc
cgggtggctgagaacgtgctgcgttcgcagggtatccgcgcctggccggt
ctgcggccgccgcggctga

mtsp39
atgagcaccatcttcgacatccgcagcctgcgactgccgaaactgtctgc
aaaggtagtggtcgtcggcgggttggtggtggtcttggcggtcgtggccg

FIG. 2 (continued)

ctgcggccggcgcgcggctctaccggaaactgactaccactaccgtggtc
gcgtatttctctgaggcgctcgcgctgtacccaggagacaaagtccagat
catgggtgtgcgggtcggttctatcgacaagatcgagccggccggcgaca
agatgcgagtcacgttgcactacagcaacaaataccaggtgccggccacg
gctaccgcgtcgatcctcaaccccagcctggtggcctcgcgcaccatcca
gctgtcaccgccgtacaccggcggcccggtcttgcaagacggcgcggtga
tcccaatcgagcgcacccaggtgcccgtcgagtgggatcagttgcgcgat
tccatcaatgggatcctccgccagctcggcccgacggagcggcagccgaa
ggggccgttcggcgacctcatcgaatcggccgcggacaacctggccggca
agggcaggcagctcaacgaaacgctgaacagtttgtcgcaggcgttgacc
gcgctgaacgagggccggggagacttcgttgcgatcacgcgaagcctggc
gctatttgtcagcgcgctctaccagaatgatcaacagttcgttgcgctca
acgaaaaccttgccgagttcaccgactggttcaccaaatccgaccatgac
ttggccgacacggtggaacggatcgacgacgttctcggcaccgtccgaaa
gttcgtgagcgacaacagatccgtgctggctgccgatgtcaacaacctcg
ccgacgcgaccactacactagtgcaacccgagccgcgggacggtctggaa
accgcgttgcacgtgttgccgacctacgccagcaacttcaacaaccttta
ctatccactgcacagctctctggtgggccagttcgtgttccccaacttcg
cgaacccaattcagctcatttgcagcgctattcaggccggcagccgactc
ggctatcaggaatccgccgagctgtgcgcgcagtacttggcaccggttct
ggacgctctcaagttcaattacttgccgttcggctcaaacccgttcagtt
cggcggccactttgcccaaggaggtggcttactccgaggagcggctccgc
ccgccgcccgggtacaaggacaccactgtcccagggatcttctcgcggga
cacaccgttttcacacggcaaccatgaaccgggctgggtcgttgcgcccg
ggatgcagggtatgcaggttcagccgtttaccgcgaacatgctcaccccg
gaatcgctggcagagctgctgggtggtccggatattgccccccgccgcc
gggaaccaacttgcccggaccgccgaatgcgtatgacgagtccaatccgt
tgccgccgccgtggtacccgcagcccgcgtccctcccggctgcgggcgcc
acaggacagccaggcccgggccagtga

mtsp40
atgaaacgcagcatgaaaagcggctccttcgcgatcggtctggcaatgat
gctcgccccgatggtggccgcgcccggtcttgcggccgcagacccggcca
cgcggccggtggattatcaacagatcaccgacgtcgtgatcgcgcgcggg
ctgtcgcagcgcggcgtgccgttctcctgggccggcggcggcatcagcgg
ccccacgcgcggcaccggtaccggcatcaacaccgtcgggttcgacgcct
ccggtttgatccagtacgcctatgccggtgccgggctaaagctgccgcgt
tcttccggccagatgtacaaggttgggcaaaaggtcctgccgcagcaagc
gcgcaagggcgacctgatcttctacggccccgaaggcacgcaaagcgtcg
cgttatacctcgggaagggccagatgctggaggtgggcgacgtcgtccag
gtttcgccggtgcgcaccaacggcatgacgccttacctggtccgggttct
cgggacccagccgacgcccgtccaacaggcgccggtccagccagcgccgg
tccagcaagcgcccgtccagcaagcgcccgtccaacaggcgcccgtccaa
caggcgccggtccaacaggcgccggtccagcaagcgcccgtccagcaagc
gcccgtccagccgcctcccttcggcaccgcgcgctcacgctaa

mtsp41
atgttcactcgccgtttcgccgcctccatggttggcaccaccttgactgc
cgctactttgggcctggccgcactcggcttcgccgggaccgccagcgcaa
gctcgaccgacgaagcgttcctcgcgcagctgcaggcggacgggatcact

FIG. 2 (continued)

ccgccgagcgcagcgcgcgccatcaaggacgcgcacgccgtctgcgacgc
cctcgacgagggtcactcggccaaagcggtcatcaaggcggtggccaagg
cgaccggtctgagcgccaagggcgccaagacgttcgccgttgacgccgcg
tcggcctactgcccgcagtacgtgacctcgagctaa

mtsp42
atggcggccatgtggcgccgcagaccgttgagctcggcgctgctgtcctt
cgggttgctgctcggcggactgcccctagcagcgccccgttggccggcg
cgactgaagaacccggcgccggccaaaccccgggtgcgccggtcgtggcg
ccgcaacagagttggaacagctgccgcgagttcatcgccgacaccagcga
aattcgcactgcacgctgcgcgacggtgtccgtccccgtcgactacgacc
aacccggtgggacacaagcgaagttggcggtgatccgcgtccccgcgacg
ggacagcgattcggagcactgctggtcaatcctgggggacccggggcgtc
ggcggtcgacatggtcgccgctatggcacccgcgatcgccgacaccgaca
ttctccgccacttcgacctggtgggcttcgacccgagaggggtcggccac
tcgacccctgcgttgcggtgtcgcaccgacgccgagttcgacgcgtaccg
gcgcgatccgatggccgactacagtccggccggtgtcacccacgtcgaac
aggtctaccggcagttggcccaggactgtgttgaccggatgggcttcagc
ttcttggccaatatcggtaccgcgtccgtcgcacgggacatggacatggt
tcgccaagcgttaggtgacgatcagatcaactacctcggatacagctacg
gcaccgagttgggcaccgcttacctggaacggttcggtactcatgtgcgg
gcgatggtcctcgacggcgctatcgatccagccgttagcccaatcgagga
aagcatcagccaaatggcgggatttcagaccgctttcaatgactacgccg
ccgactgcgcccgctcgccggcctgccctctgggcaccgactcggcccag
tgggtcaaccgctaccacgccctggttgacccgctggtgcagaagccggg
taagacgtcggatccacgtggcctgagctacgccgacgcgacgacgggca
ccatcaacgcgctgtacagccctcagcgctggaagtacctgaccagtggt
ctgctggggctgcagcgcggcagcgacgccggcgacttgctggtgcttgc
cgacgactatgacggccgggatgcagacgggcactacagcaacgaccagg
acgcgttcaacgcggtccggtgcgtcgatgcgcccacaccggccgatcca
gcggcctgggtggccgccgaccaacggatccgtcaggtcgccccgttcct
tagctacgggcagttcaccggatccgccccccgcgatctgtgcgcgctgt
ggccggtgccggcaacgtcgacgccgcaccccgcggcgccggccggggct
ggcaaggtcgtcgtggtgtccaccacccacgacccggccactccgtatca
gtccggggtagacctggcccgccagctgggcgcaccgctgatcaccttcg
acggcacccaacacactgcggtgttcgatggcaaccagtgtgtggactct
gcggtgatgcactattttctcgacgggaccttgccgccgacgagtctgcg
gtgcgcgccctga

mtsp43
atgaagacaggcaccgcgacgacgcggcgcaggctgttggcagtactgat
cgccctcgcgttgccgggggccgccgttgcgctgctggccgaaccatcag
cgaccggcgcgtcggacccgtgcgcggccagcgaagtggcgaggacggtc
ggttcggtcgccaagtcgatgggcgactacctggattcacacccagagac
caaccaggtgatgaccgcggtcttgcagcagcaggtagggccggggtcgg
tcgcatcgctgaaggcccatttcgaggcgaatcccaaggtcgcatcggat
ctgcacgcgctttcgcaaccgctgaccgatctttcgactcggtgctcgct
gccgatcagcggcctgcaggcgatcggtttgatgcaggcggtgcagggcg
cccgccggtag

FIG. 2 (continued)

mtsp44
atgtctcggctgagttccatcctgcgtgccggcgcggcatttctggttct
cggcatcgccgctgcgacatttccacaaagcgcggcagccgactccacgg
aagactttccaatacctcgccggatgatcgcaaccacctgcgacgccgaa
caatatctggcggcggtgcgggataccagtccggtgtactaccagcggta
catgatcgacttcaacaaccatgcaaaccttcagcaagcgacgatcaaca
aggcgcactggttcttctcgctgtcaccggcggagcgccgagactactcc
gaacacttttacaatggcgatccgctgacgtttgcctgggtcaatcacat
gaaaatcttcttcaacaacaagggcgtcgtcgctaaagggaccgaggtgt
gcaatggatacccagccggcgacatgtcggtgtggaactgggcctaa

mtsp45
gtgaccaagcgcacaataactcccatgacgtcgatgggtgatctcttgggacctgagcca
atcctgttgcctggcgacagcgacgccgaagcggagctgcttgccaacgaaagtccgagc
atcgtcgcggccgcgcatccgtcggcgtcggtcgcctgggcggtgctcgccgaaggggcg
ctggccgacgacaagaccgtcacggcctacgcatacgcgcgtaccgggtaccaccgcggc
ctggaccagctgcgccgccatggctggaagggcttcggcccggtgccgtattcccaccag
cccaaccggggtttcctacggtgtgtggcggcgctggcgcgcgccgcagccgctatcggc
gagaccgacgagtatggacgctgcctggatctgcttgacgactgtgaccccgcggcccgt
ccggcgcttgggctc

mtsp46
gtgatcatccctgacatcaatctgctgctctacgcggtcatcaccggattcccgcagcac
cggcgcgcgcatgcgtggtggcaagacaccgtcaacggccacacccgtatcgggctgacg
tatccggcgttgttcgggttcctacggatcgccaccagtgcccgcgtgctcgccgcgcca
ctgccaaccgcggatgcgatcgcctatgtgcgcgagtggctttcgcagccgaacgtggac
ctactcacggcgggtccgcgccacctggacatcgcgttgggcctgctcgacaagctcggc
acagccagccacctaaccaccgatgtgcaactggccgcctacggcatcgaatacgacgcc
gagatccattccagtgacaccgactttgcccgattcgccgatctgaagtggaccgacccg
ttgcgcgaa

mtsp47
ttgactgatccgcgccacaccgttcgaatcgctgtcggagctaccgcgctcggcgtgtcg
gcactcggggcaactctgccggcctgctccgcacacagcgggccgggttctccccccagt
gcgccgtcagctcccgcggccgcgaccgtcatggtagagggacatacgcacacaatttcc
ggagtggtcgagtgccgcacctcgccagcggtaaggacggcgacgccgtcggagtcgggg
actcaaactacacgggttaacgcacacgacgattcggcctcggtgacactgtccctgtcc
gactccacgcccccagacgtcaatggttttggtatctcccttaaaatcggaagcgtcgac
taccagatgccctaccagccggttcagtccccaactcaggtcgaagcgaccaggcagggc
aagagttacacactgaccgggacgggtcacgcggtgatcccgggccaaaccggcatgcgt
gagctgccgttcggggtacatgtaacctgtccg

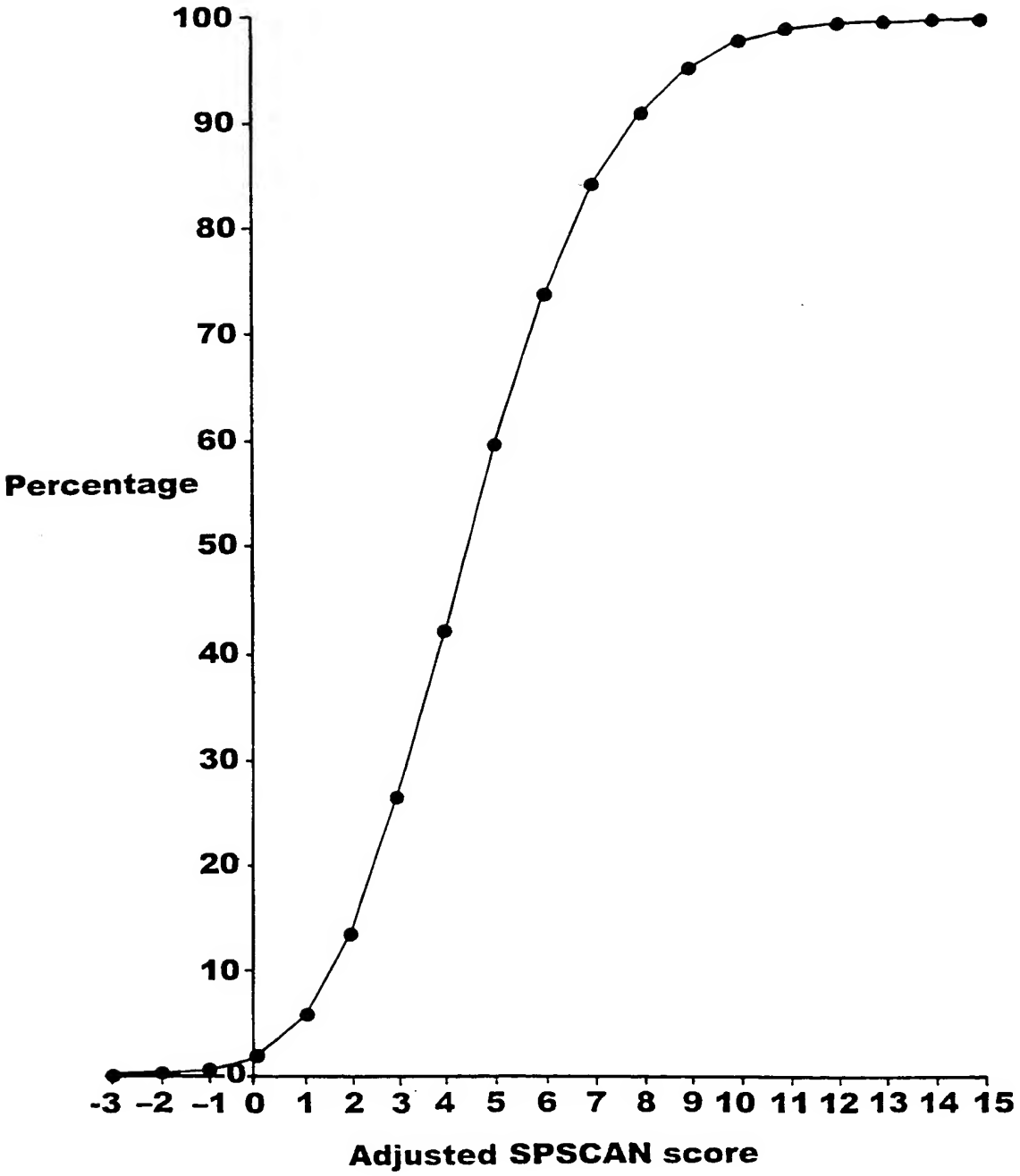
100
90
80
70
60
Percentage
50
40
30
20
10
0
-3 -2 -1 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Adjusted SPSCAN score

FIG. 3A

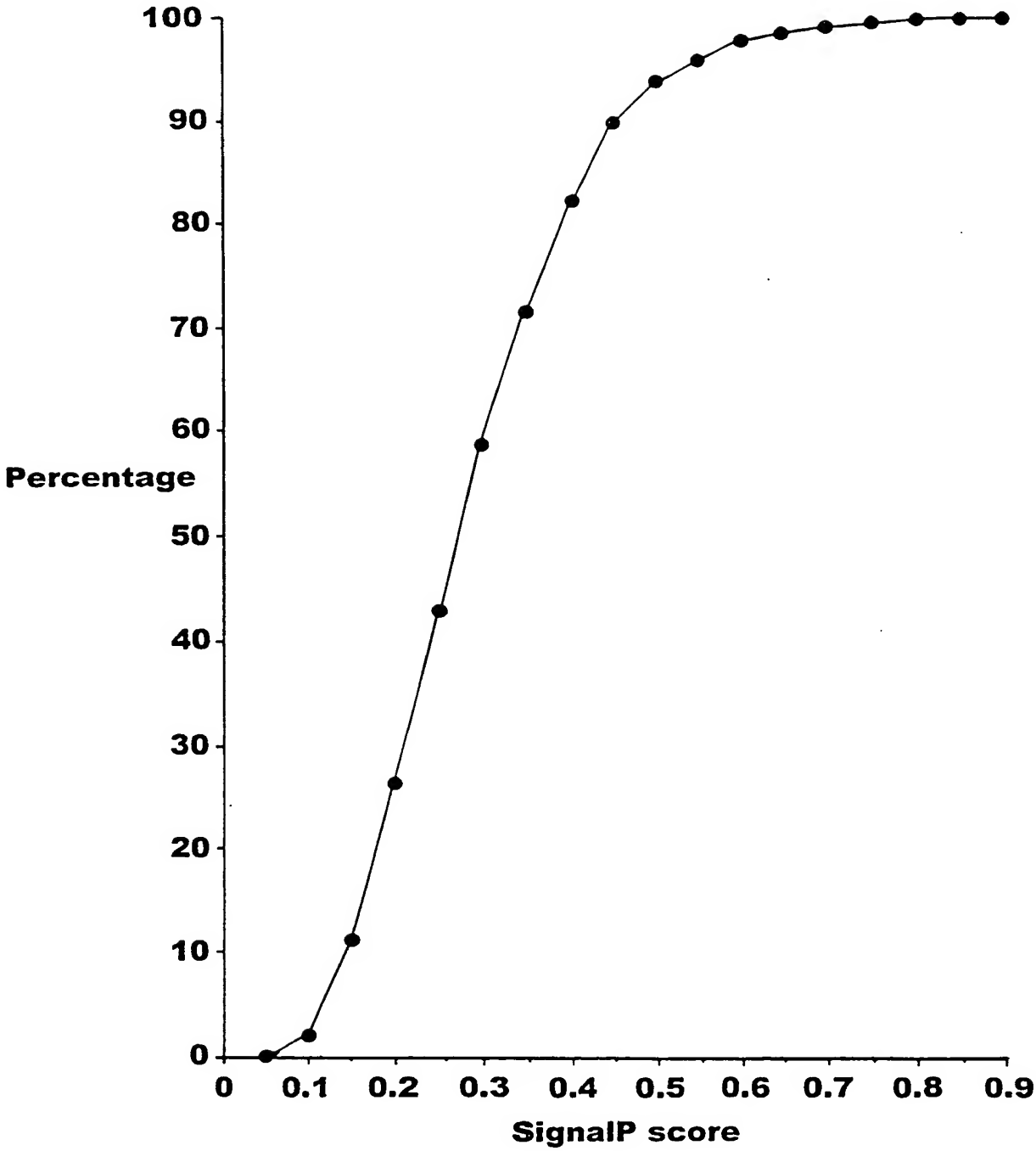
100
90
80
70
60
Percentage
50
40
30
20
10
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
SignalP score

FIG. 3B

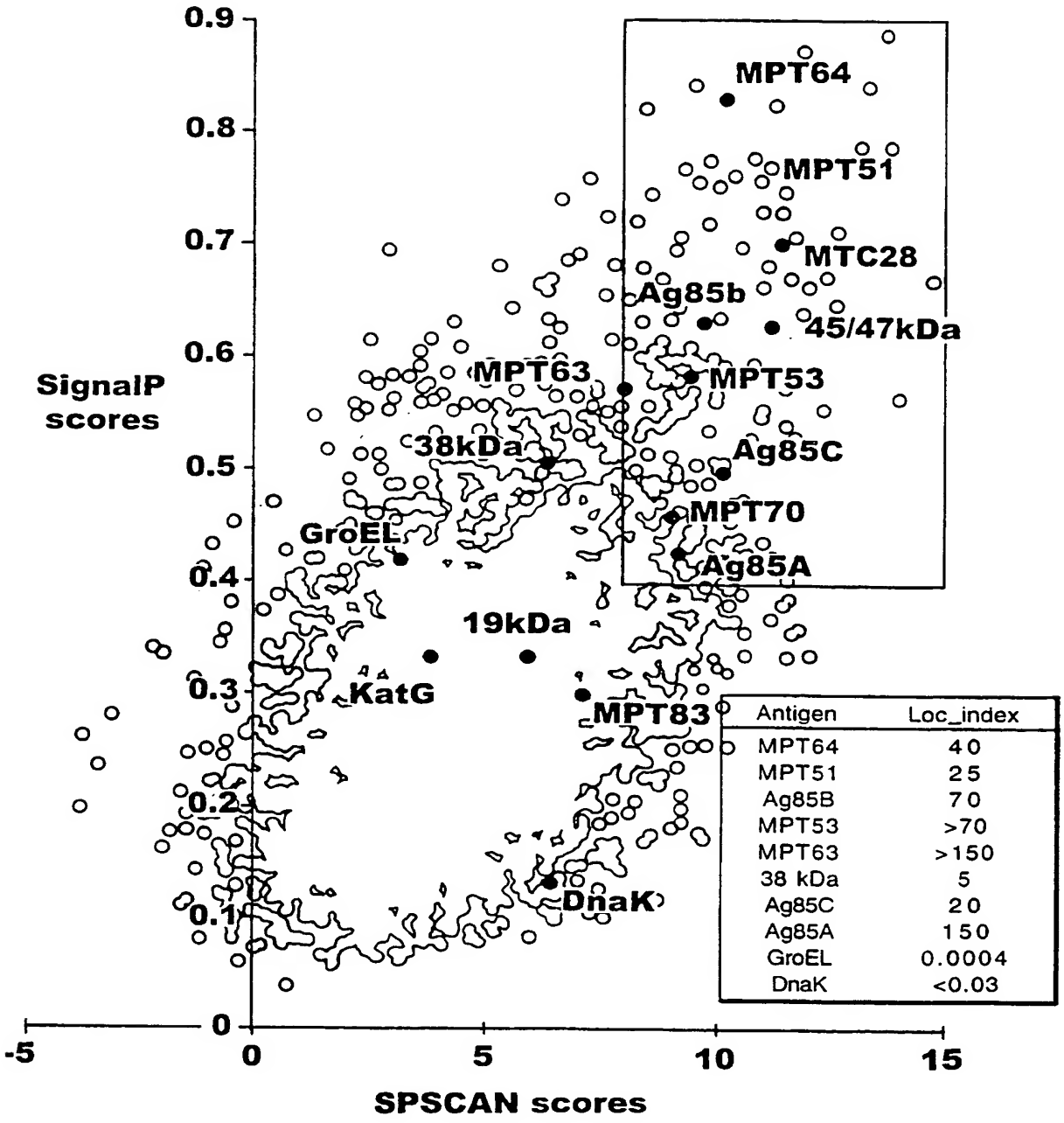
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
SignalP scores
-5
0
5
10
15
SPSCAN scores
MPT64
MPT51
MTC28
Ag85b
45/47kDa
MPT63
MPT53
38kDa
Ag85C
MPT70
GroEL
Ag85A
19kDa
KatG
MPT83
DnaK
Antigen | Loc_index
MPT64 | 40
MPT51 | 25
Ag85B | 70
MPT53 | >70
MPT63 | >150
38 kDa | 5
Ag85C | 20
Ag85A | 150
GroEL | 0.0004
DnaK | <0.03

FIG. 3C

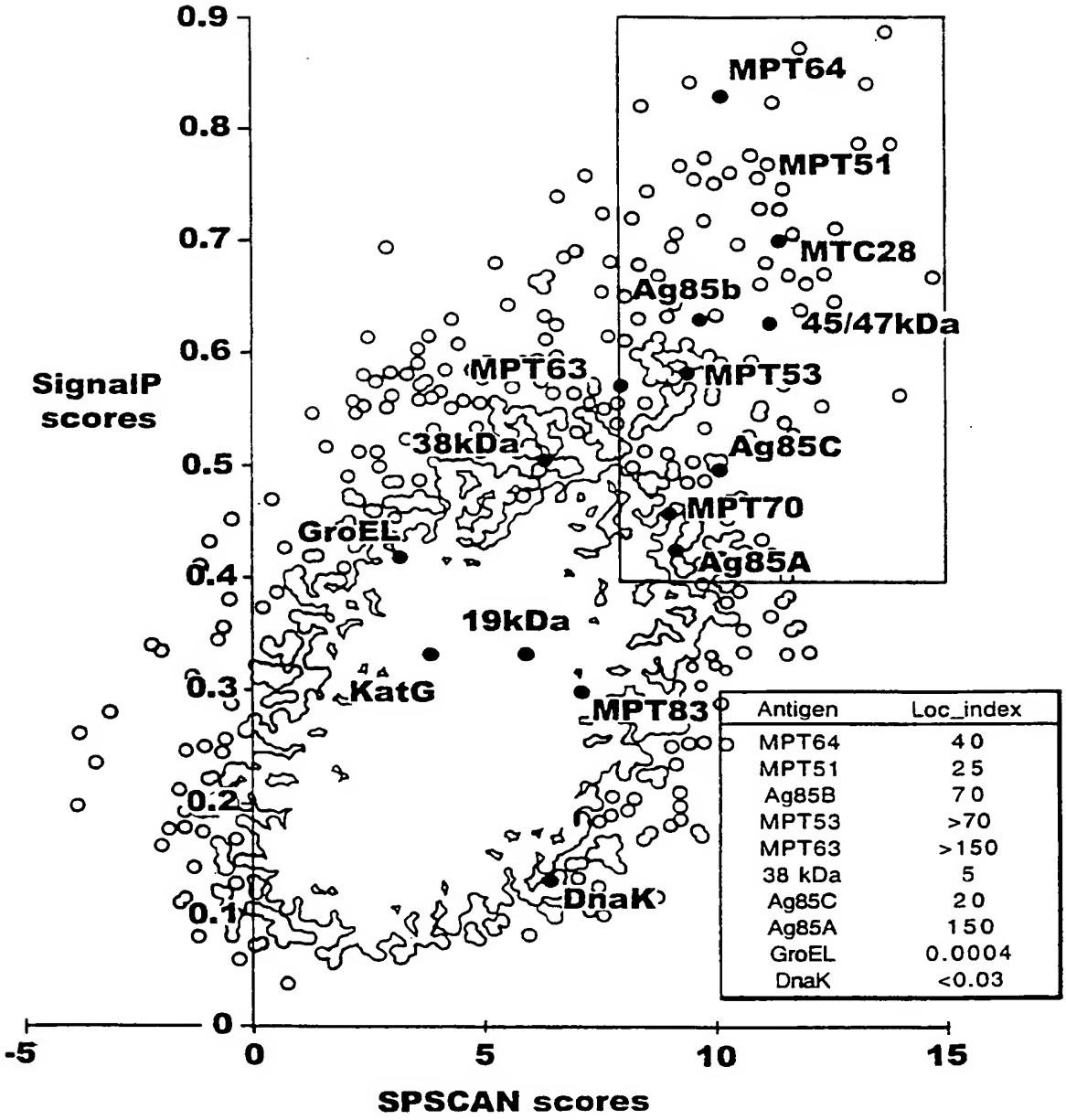
SignalP scores
SPSCAN scores
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-5
0
5
10
15
MPT64
MPT51
MTC28
Ag85b
45/47kDa
MPT63
MPT53
38kDa
Ag85C
MPT70
GroEL
Ag85A
19kDa
KatG
MPT83
DnaK
Antigen Loc_index
MPT64 40
MPT51 25
Ag85B 70
MPT53 >70
MPT63 >150
38 kDa 5
Ag85C 20
Ag85A 150
GroEL 0.0004
DnaK <0.03

FIG. 4

# INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/12197

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(7) :Please See Extra Sheet.
US CL :Please See Extra Sheet.
According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 424/185.1, 234.1, 248.1; 435/69.1, 71.1, 91.1, 253.1; 530/300, 350; 536/22.1, 23.1, 23.7

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

FIOSIS, CABA, CAPLUS, EMBASE, LIFESCI, MEDLINE, SCISEARCH, USPATFULL, JAPIO

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| Y | MANCA, C. et al. MTC28, a novel 28-kilodalton proline-rich secreted antigen specific for the Mycobacterium tuberculosis complex. Infection and Immunity. December 1997, Vol. 65, No. 12, pages 4951-4957, entire reference. | 1-20 |
| Y | MANCA, C. et al. Molecular cloning, purification, and serological characterization of MPT63, a novel antigen secreted by Mycobacterium tuberculosis. Infection and Immunity. January 1997, Vol. 65, No. 1, pages 16-23, entire reference. | 1-20 |

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

| | | | |
|---|---|---|---|
| * | Special categories of cited documents: | "T" | later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention |
| "A" | document defining the general state of the art which is not considered to be of particular relevance | "X" | document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone |
| "E" | earlier document published on or after the international filing date | "Y" | document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art |
| "L" | document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) | "&" | document member of the same patent family |
| "O" | document referring to an oral disclosure, use, exhibition or other means | | |
| "P" | document published prior to the international filing date but later than the priority date claimed | | |

| | |
|---|---|
| Date of the actual completion of the international search<br>06 SEPTEMBER 2000 | Date of mailing of the international search report<br>22 SEP 2000 |
| Name and mailing address of the ISA/US<br>Commissioner of Patents and Trademarks<br>Box PCT<br>Washington, D.C. 20231<br>Facsimile No. (703) 305-3230 | Authorized officer<br>RODNEY P. SWARTZ, PH.D.<br>Telephone No. (703)308-0196 |

Form PCT/ISA/210 (second sheet) (July 1998)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/12197

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| Y | GENNARO, M.L. et al. Gene cloning and purification of proteins secreted by Mycobacterium tuberculosis. Journal of Cellular Biochemistry Supplement. 1995, Abstract No. B3-112, page 68, entire abstract. | 1-20 |
| Y | ANDERSEN, P. et al. Proteins released from Mycobacterium tuberculosis during growth. Infection and Immunity. June 1991, Vol. 59, No. 6, pages 1905-1910, entire reference. | 1-20 |
| Y | BORREMANS, M., et al. Cloning, sequence determination, and expression of a 32-kilodalton-protein gene of Mycobacterium tuberculosis. Infection and Immunity. October 1989, Vol. 57, No. 10, pages 3123-3130, entire reference. | 1-20 |
| Y | CONTENT, J. et al. The genes coding for the antigen 85 complexes of Mycobacterium tuberculosis and Mycobacterium bovis BCG are members of a gene family: Cloning, sequence determination, and genomic organization of the gene coding for antigen 85-c of Mycobacterium tuberculosis. Infection and Immunity. September 1991, Vol. 59, No. 9, pages 3205-3212, entire reference. | 1-20 |
| Y | HORWITZ, M.A. et al. Protective immunity against tuberculosis induced by vaccination with major extracellular proteins of Mycobacterium tuberculosis. Proceedings of the National Academy of Sciences, USA. February 1995, Vol. 92, pages 1530-1534, entire reference. | 1-20 |
| Y | ROBERTS, A.D. et al. Characteristics of protective immunity engendered by vaccination of mice with purified culture filtrate protein antigens of Mycobacterium tuberculosis. Immunology. 1995, Vol. 85, pages 502-508, entire reference. | 1-20 |
| Y | MATSUMOTO, S. et al. Cloning and sequencing of a unique antigen MPT70 from Mycobacterium tuberculosis H37Rv and expression in BCG using E. coli-Mycobacteria shuttle vector. Scandinavian Journal of Immunology. 1995, Vol. 41, pages 281-287, entire reference. | 1-20 |
| Y | LAQUEYRERIE, A. et al. Cloning, sequencing, and expression of the apa gene coding for the Mycobacterium tuberculosis 45/47-kilodalton secreted antigen complex. Infection and Immunity. October 1995, Vol. 63, No. 10, pages 4003-4010, entire reference. | 1-20 |

# INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/12197

**Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)**

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

**Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-20

**Remark on Protest**

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

# INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/12197

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (7):

A61K 38/00, 39/00, 39/02, 39/04;C07H 21/02, 21/04; C07K 1/00, 14/00; C12N 1/12, 1/20; C12P 19/34, 21/04, 21/06

A. CLASSIFICATION OF SUBJECT MATTER:
US CL :

424/185.1, 234.1, 248.1; 435/69.1, 71.1, 91.1, 253.1; 530/300, 350; 536/22.1, 23.1, 23.7

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claims 1-20, drawn to DNA, vector, transformed cells, polypeptides, and a first method of use for diagnosis in vivo.
Group II, claims 21-24, drawn to second method of use for diagnosis in vitro using cells.
Group III, claims 25-28, drawn to third method of use for diagnosis in vitro using antibodies.
Group IV, claims 29-36, drawn to fourth method of use for vaccination.

The inventions listed as Groups I-IV do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the DNA is a single inventive concept, and as such the corresponding vectors, transformed cells, polypeptides encoded by the DNA, and a first method of use of the polypeptides are included in the first invention. The remaining Groups II-IV are additional methods of use, not requiring the DNA, but can utilize isolated and purified polypeptides obtained from *M. tuberculosis*.

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :
A61K 38/00, 39/00, 39/02, 39/04, C07H 21/02, 21/04, C07K 1/00, 14/00, C12N 1/12, 1/20, C12P 19/34, 21/04, 21/06

A1

(11) International Publication Number: WO 00/66143

(43) International Publication Date: 9 November 2000 (09.11.00)

(21) International Application Number: PCT/US00/12197

(22) International Filing Date: 4 May 2000 (04.05.00)

(30) Priority Data:

| | | |
|---|---|---|
| 60/132,479 | 4 May 1999 (04.05.99) | US |
| 60/132,503 | 4 May 1999 (04.05.99) | US |

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications

| | |
|---|---|
| US | 60/132,503 (CON) |
| Filed on | 4 May 1999 (04.05.99) |
| US | 60/132,479 (CON) |
| Filed on | 4 May 1999 (04.05.99) |

(71) Applicant *(for all designated States except US)*: THE PUBLIC HEALTH RESEARCH INSTITUTE OF THE CITY OF NEW YORK, INC. [US/US]; 455 First Avenue, New York, NY 10016 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: GENNARO, Maria, L. [US/US]; – (US). GOMEZ, Manuel, J. [US/US]; – (US).

(74) Agent: HONE, William, J.; Fish & Richardson P.C., Suite 2800, 45 Rockefeller Plaza, New York, NY 10111 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published**

*With international search report.*

*Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.*

(54) Title: SECRETED PROTEINS OF *MYCOBACTERIUM TUBERCULOSIS* AND THEIR USE AS VACCINES AND DIAGNOSTIC REAGENTS

(57) Abstract

The invention provides *mycobacterium tuberculosis* polypeptides and genes encoding them for use in diagnostic and prophylactic methodologies.